

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

CELL BREAKOUT

*How loose DNA from unstable
chromosomes helps to drive
cancer metastasis* **PAGE 467**

SCIENTIFIC METHOD

REPETITIVE STRAIN

*Triangulate evidence to
verify experimental results*

PAGE 399

APPLIED PHYSICS

IMAGES OUT OF THIN AIR

*Laser turns trapped particles
into 3D moving display*

PAGES 408 & 406

BIOTECHNOLOGY

LIPSTIC TRACES

*Enzymatic labelling tracks
transient cell-cell contacts*

PAGE 414 & 496

NATURE.COM/NATURE

25 January 2018 £10

Vol. 553, No. 7689



THIS WEEK

EDITORIALS

POLITICS Reflections on the first twelve months of President Trump **p.380**

WORLD VIEW Herbicide row shows value of independent advice **p.381**



CLIMATE Rising temperatures linked to antelope die-off **p.383**

Science has a gambling problem

Researchers and government agencies pay too little attention to pathological gambling. This must change.

Pathological gambling is thought to affect as many people as autism and schizophrenia. It disrupts employment, relationships and health, and places an enormous burden on the state. It is the only behavioural addiction formally recognized by the American Psychiatric Association, appearing in the fifth edition of the organization's *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* in 2013. And what is the contribution of science to this pressing debate? A review last year of all research literature looked for well-designed studies conducted in real gambling environments with real gamblers (R. Ladouceur *et al. Addiction Res. Theory* **25**, 225–235; 2017). It found just 29. In total.

No one is calling for a prohibition on gambling, a legitimate leisure pursuit. Most people can enjoy the occasional flutter without harm. But how can research help the unfortunate minority who cross to gambling's dark side? Or provide enough evidence for society to control an industry that gains more from compulsive than from occasional gambling?

The first essential step is to look beyond the glitz. The media lapped up a story in 2014 of how Saudi heiress Nora Al-Daher, also the wife of an Omani government minister, tried to sue the Ritz casino in London for allowing her to run up a £2-million (US\$2.8-million) gambling debt. She argued that the casino took advantage of her addiction, raising her agreed cheque limit during a high-rolling binge when she couldn't have been held responsible for her actions. The judge wasn't impressed, and ruled against her. The court of public opinion had little sympathy, too.

For every Saudi heiress, there are millions of desperate people who slip from normal behaviour to abuse and addiction. They need help — not pity or scorn — and a solid evidence base to build a system to protect them.

The second step is to forget the popular image of horses and roulette tables. Most people with a gambling addiction play online. And they do 'play': the distinction between online gambling and online gaming is being eroded, as the two multinational industries exchange tips to draw people in and keep them playing for longer. (Indeed, psychiatrists are looking again at whether they need a new diagnosis of compulsive playing of computer games. The World Health Organization plans to introduce 'gaming disorder' into its International Classification of Diseases this year.)

The world of gambling research is too small and underfunded. The paucity of data available to inform policymakers and the medical profession is shocking. Much more needs to be understood about the elements of diverse online and offline gambling activities — for example, display strategies on screens that mislead users on the chances of winning — and the epidemiology of who is most vulnerable and so most likely to be seduced by the lure of addiction.

Many countries have adopted a formal, non-binding Responsible Gambling Strategy — based on the original 2004 Reno Model in the United States, which developed voluntary guidelines for the industry — to address problematic gambling. These strategies aim mostly to devise actions to protect those who are vulnerable to gambling; these include,

for instance, advance agreements with casinos to limit the size of bets for people known to have an addiction, or to ban them from a gambling establishment for a fixed period. There is little empirical evidence as to whether such strategies work.

In some countries, the strategies have also spawned the creation of funds to support research, such as the US National Center for Responsible Gaming (NCRG), which has distributed \$27 million to researchers since it was created in 1996. The NCRG has processes for handling bottom-up research applications according to fair, peer-reviewed procedures, but because it is financed by the industry — casino companies, equipment manufacturers and the like — some fear that research

“There is an unquestionable need for vigilance.”

agendas could be distorted. Some social scientists worry that distortion is already visible because so much of the funding goes to support research into the behaviour of individuals who gamble, as opposed to the role of industry and society. They argue that this

inappropriately shifts responsibility from the industry — which wants to minimize regulation — to individuals. They fear a parallel with how the tobacco industry managed to distort research into the dangers of smoking. According to a 2014 report from ethnologist Rebecca Cassidy of Goldsmiths, University of London, many researchers feel uncomfortable accepting support from the NCRG's UK equivalent, GambleAware, for this reason (R. Cassidy *Int. Gambl. Stud.* **14**, 345–353; 2014).

To be fair, the NCRG does fund some research into public-health issues around gambling. But there is an unquestionable need for vigilance. It isn't appropriate for research related to a major social and public-health problem to be so heavily dependent on the very industry that enables it. Instead, governments need to design and support their own research programmes to ensure that the appropriate range of reliable evidence is generated to inform policymaking and health organizations.

Which agency should take the lead? Gambling doesn't attack any particular organ in the body in the way that smoking attacks lungs, and alcohol the liver. So government health-research agencies that have experience tackling substance abuse — for example, the National Institute on Drug Abuse and the National Institute on Alcohol Abuse and Alcoholism in the United States — have not been particularly involved. Problem gambling does cause health problems, however, not only through self-neglect, but perhaps also as a result of its extraordinarily high level of comorbidity with other psychiatric conditions; psychiatrists have started to investigate whether gambling in itself could precipitate a psychotic incident in someone who was previously in a subclinical state.

Irrespective of possible blurring of responsibilities, each government needs to make a call and assign the problem of pathological gambling to an appropriate agency or ministry. And more scientists must respond to the very real need to assess and understand the implications, in the same way as those before them have done so admirably on the abuse of alcohol, tobacco and drugs. ■

The Great Flu

It's been a century since the world's worst influenza pandemic — could it happen again?

One hundred years ago this month, the 1918 influenza virus was just starting to spread. It would become the greatest public-health crisis of the twentieth century, claiming some 50 million to 100 million lives. The centenary has raised questions over whether such a severe flu pandemic could happen today, and whether the world is prepared.

There are few data points to go on — flu pandemics happen only three or four times a century — but one risk is certainly higher: 7.6 billion people share the planet in 2018, up from 1.9 billion in 1918. Feeding all those extra people has also meant a huge rise in livestock numbers, intensive farming and the numbers of animals being transported around the world. Scientists say that the genetic mixing and evolution of animal flu viruses is thus being amplified, increasing the chance of viruses gaining the potential to jump to humans and, if they can spread easily between people, causing a human pandemic. Our just-in-time global production systems and service economies are also exquisitely vulnerable to the quickly cascading disruption that a severe pandemic would cause.

The case-fatality rate in the 1918 pandemic was around 2.5% (compared with less than 0.1% in other flu pandemics), and a comparable or worse rate in a future pandemic cannot be discounted. There are two hypotheses to explain the 1918 strain's high lethality: cytokine storms and secondary bacterial infection. (In a cytokine storm, the body's immune system overreacts, causing tissue and organ damage, and even death.)

But in an intriguing 2008 paper (D. M. Morens, J. K. Taubenberger and A. S. Fauci *J. Infect. Dis.* **198**, 962–970; 2008), researchers went through data from almost 8,500 post-mortem records from the 1918–19 pandemic and discovered what doctors knew at the time, but which was subsequently forgotten — that most people in the pandemic probably

died of secondary pneumonia from common bacterial pathogens.

Were this latter pattern to dominate in any severe future pandemic, the availability of antibiotics, which didn't exist in 1918, would dent death rates, provided that sufficient stockpiles were available. More broadly, the importance of robust public-health systems and surge capacity in hospitals as a basic bulwark against epidemic and pandemic threats of all kinds cannot be overstated. Yet health systems remain weak in many countries.

Speak to scientists, and they all agree on what must be the number one research goal for effective mitigation of any future pandemic: a universal flu vaccine. At present, the seasonal flu vaccine usually has to be updated every year or so to match the circulating virus strains — which are continually evolving — and these vaccines provide no protection against an altogether new pandemic subtype.

The 2009 swine flu pandemic showed that it takes months to start producing a vaccine against a pandemic flu virus. In many countries, substantial amounts of vaccine arrived only after the first wave of infection had already passed. Fortunately, the 2009 virus was relatively mild.

A universal vaccine, ideally offering lifelong protection against all flu subtypes, would improve the effectiveness of seasonal flu vaccines and offer protection against novel pandemic threats. Even a vaccine that is around 75% effective in preventing disease symptoms would be a huge public advance, scientists reckon (C. I. Paules *et al. Immunity* **47**, 599–603; 2017).

A major international workshop on developing such a vaccine was held last year in Rockville, Maryland, and identified many research gaps — including the complexity of the immune response to infection and vaccination — and a road map for addressing them. Yet the United States, one of the largest flu-research funders, last year invested just US\$75 million on universal flu vaccine research and development.

Whether the world will again ever see the likes of the 1918–19 flu pandemic cannot be reliably predicted, but given the stakes, it is best for society, as a whole, to plan for worst-case scenarios. And advocates rightly argue that the research and development of a universal flu vaccine — ultimately the only effective defence against future pandemics — merits a programme equivalent in scale to the Manhattan Project. ■

Annual report

Donald Trump has been in office for a year and the effects on science have been as bad as feared.

After a year of President Trump, scientists in the United States are doing their best in difficult circumstances, and *Nature* applauds them for it. It's increasingly clear that Trump has been just as bad for many aspects of science as we and others feared. Most crucially, the role of science and scientific advice in public life has been repeatedly undermined.

Writing after his election victory in November 2016, this journal tried to look on the bright side and suggested that Trump could yet “leave behind his damaging and unpopular attitudes and embrace reality, rationality and evidence” (*Nature* <http://doi.org/bs57>; 2016).

How wrong we were to be optimistic. After 12 months in office, Trump's impact on science can be neatly divided into two categories: bad things that people expected, and bad things that they didn't. The long list of items in the first category includes the US withdrawal from the Paris climate agreement, regulatory rollback across government (environmental agencies in particular) and the now record-breaking failure to appoint a science adviser. His administration has cut off funds to organizations abroad that promote public health but mention

abortion, weakened restrictions under the Toxic Substances Control Act and censored the use by government agencies of phrases such as “evidence-based” and “climate change”. Advisory groups, including one on HIV/Aids, have been disbanded, and scientists with Environmental Protection Agency grants have been banned from serving on the agency's advisory boards.

Turning to the second category, Trump's campaign rhetoric promised a travel ban on Muslims, but the full, chilling and chaotic details turned out to be much worse, and more divisive and disruptive, than even avowed opponents might have dared to suggest. Scientific organizations queued up to complain about the likely loss of talent.

There are also some bad things that critics expected Trump to do, but that have yet to come to pass. Budgets at key science and health agencies remain largely unmolested (although this is largely thanks to resistance in Congress to pledged cuts); bans on research using fetal tissue and embryonic stem cells have not emerged; and Obama-era programmes including the Precision Medicine Initiative remain in place for now.

One good thing has happened: Trump has triggered a surge of political activity by scientists motivated to oppose him. (And, of course, nations elsewhere, from China to France, are already stepping in to offer opportunities as US leadership slips.) Those who cherish the values of science should keep fighting. Scientists and politicians must continue to challenge the president's actions and seek to hold him to account. ■



Don't attack science agencies for political gain

Eroding trust in regulatory agencies will not improve democratic accountability, warns Bernhard Url.

The job of the European Food Safety Authority (EFSA) is to assess what might make food unsafe. That's hard enough. It is even harder when the agency is at the centre of a public debate that goes far beyond science.

This has happened with artificial sweeteners, genetically modified (GM) organisms and glyphosate, the world's most ubiquitous herbicide. When questions about a society's values are thrust onto scientific agencies rather than elected officials, scientific assessment suffers.

The glyphosate controversy began in earnest two-and-a-half years ago, when EFSA and experts designated by European Union members concluded that the product is unlikely to be carcinogenic. In late 2017, the European Commission renewed a licence allowing the herbicide's sale. EFSA's conclusion contradicted that of the International Agency for Research on Cancer (IARC), which had classified the chemical as "probably carcinogenic" months earlier, bringing its own share of controversy.

That the agencies reached different conclusions is not surprising: each considered different bodies of scientific evidence and methodologies. Other independent assessments — by the European Chemicals Agency and regulatory bodies in the United States, Canada, Japan and Australia — agreed with EFSA. So did an expert body on pesticide residues convened by the Food and Agriculture Organization of the United Nations and the World Health Organization.

Even so, the divergence between EFSA's conclusion and the IARC's has been debated by legislators from Brussels to Berlin and beyond. We have seen scare stories about trace levels of glyphosate residues in German beer or Italian pasta — but these fail to mention that observed amounts of herbicide residues would pose risks only if a person consumed roughly 1,000 litres of beer or their body's weight in dry pasta in one day.

Why the frenzy? Agencies that find low risk of regulated products are often accused of undue industry influence. We at EFSA believe that some campaigners are unwilling to accept any evidence that certain regulated substances are safe, and will tout weak scientific studies showing the opposite. The same groups applauded EFSA for reviews on other pesticides, such as neonicotinoids, that it deemed dangerous.

It seems to us that some campaigners contest the science of safety assessments in pursuit of greater political arguments. These arguments deserve airing — but they belong with policymakers.

In the past two years, EFSA has faced multiple allegations over its evaluation of glyphosate. The most pernicious of these is that the agency violated good scientific practice by plagiarizing information from industry. It is true that the document in question, the Renewal Assessment Report produced by German authorities, includes a section summarizing published toxicology literature that contains text compiled

by a committee of some 20 companies, including glyphosate's original manufacturer, Monsanto. But this is standard practice, and EFSA peer-review panels vetted the material that appeared.

The section brought forward as allegedly copied from industry also highlights concerns over products that contain glyphosate. In fact, it was used to support a recommendation by EFSA in November 2015 to further evaluate the safety of plant-protection products containing glyphosate. This section was made publicly available for comment in 2014, but complaints of copied text by regulatory agencies came in late 2017, after other complaints were raised about Monsanto's possible influence over published scientific literature.

So, when campaigners allege that EFSA did not follow due scientific process when assessing glyphosate, we believe that they are really railing against bigger issues: the role of modern agricultural practices and multinational biotech firms in our food supply.

A broader societal discussion about these issues is essential, but it won't be achieved by picking on regulatory science. It is the role of politicians to represent the values, needs and expectations of their constituents through democratic processes. This is outside the responsibility of organizations such as EFSA, which were created to advise EU policymakers on scientific matters.

Three changes would help elected officials and regulatory agencies to do their separate jobs. First, questions about societal values should be framed ahead of and outside scientific work. The EU must equip itself with a legal and regulatory framework for food production that accounts for citizens' opinions on intensive agriculture, pesticide

use, GM organisms and other biotechnology, and the importance of biodiversity. This will provide a forum for open, honest debate.

Second, regulatory and legal guidelines should be drawn up to govern how regulatory bodies interact with industry and handle transparency of the data that they use.

Finally, politicians need to decide whether they are willing to allow risk assessment of regulated products, such as glyphosate and food additives, to continue to be based on safety studies commissioned and paid for by the industry, as has been the case for decades. If so, politicians must have the courage to support the regulatory bodies charged with implementing these rules. If not, they must find funding for these studies elsewhere. Only once these steps have been taken will regulatory agencies be free from allegations of bias when their scientific conclusions are at odds with the political agenda of one interest group or another. ■

AGENCIES THAT FIND
LOW RISK
OF REGULATED
PRODUCTS
ARE OFTEN
ACCUSED
OF UNDUE INDUSTRY
INFLUENCE.

Bernhard Url is executive director of the European Food Safety Authority in Parma, Italy.
e-mail: bernhard.url@efsa.europa.eu

SEVEN DAYS

The news in brief

POLITICS

US shutdown

The US government shut down on 20 January, after lawmakers in Congress failed to agree on legislation to fund the government before a stopgap budget measure expired. Many federal employees, including those at science agencies, were ordered to stop working, and major research funders such as the US National Institutes of Health prepared to stop processing grants. The event — which began on Trump's first anniversary in office — ended on 22 January after politicians approved a short-term funding bill that expires on 8 February. See page 389 for more.

EVENTS

CRISPR patent

The European Patent Office has revoked a key patent on CRISPR–Cas9 genome editing held by the Broad Institute in Cambridge, Massachusetts. The decision, announced on 17 January, hinged on a procedural issue: an inventor listed on the Broad's initial patent application was eventually dropped from the application without written permission from that inventor. The ruling could affect some of the other European CRISPR–Cas9 patents held by the institute. The Broad has said that it will appeal against the decision.

Dengue vaccine

Drug maker Sanofi Pasteur will refund the Philippines government for US\$28 million of unused dengue vaccine after the nation suspended its use last year, the two parties announced on 15 January. The Philippines halted its immunization programme against the tropical virus — the world's first — after 14 children who had received the vaccine

died, some with severe dengue symptoms. No causal link has been proved, but Sanofi, which is headquartered in Paris, disclosed last November that the vaccine could make dengue infections worse in those who became infected with the virus for the first time after receiving the vaccine. The company said that the refund was unrelated to safety issues.

Harassment data

The US House of Representatives' science committee has asked the Government Accountability Office (GAO) to provide it with data on sexual harassment involving federally funded researchers. The committee notes that sexual harassment

has a significant negative impact on women researchers, driving some out of science altogether. In a letter sent to the GAO on 18 January, the committee asked for information on cases of and policies relating to sexual assault and harassment at the US National Institutes of Health, National Science Foundation, Department of Agriculture, Department of Energy and NASA.

POLICY

Human-subject rule

The US government has postponed updates to its policy governing research on human subjects, known as the Common Rule. The

burning biomass, which could have adverse environmental effects. The policies are not yet legally binding: the Parliament will now need to negotiate the plan with national governments, which could attempt to lower the targets. The EU accounts for about 10% of global greenhouse-gas emissions. Around 17% of the energy consumed in the region comes from renewable sources (pictured, wind turbine in France).

Europe backs bigger clean-energy targets

European lawmakers have moved to raise the European Union's renewable-energy targets. In a vote on 17 January, the European Parliament decided that by 2030, 35% of energy consumed in the EU should be from renewable sources such as wind and solar power — but not from nuclear. The existing goal is 27%. Critics say that raising clean-energy targets might prompt countries to produce more electricity by

changes were supposed to go into effect on 19 January, but a mid-January notice from 16 government agencies announced that institutions conducting such research now have until 19 July to comply with the new rules. The changes include alterations to patient consent forms, streamlined ethics reviews of proposed experiments and greater transparency requirements for study methodologies and results. Only institutions with federal grants are required to comply with the Common Rule.

Electric fishing

The European Parliament voted on 16 January to ban a controversial electric fishing



LOIC VENANCE/AFP/GETTY

technique in European Union waters. 'Pulse trawling' uses bursts of electric currents to coax flatfish out of the seabed and into nets, and is currently used mainly by Dutch vessels in the North Sea. Some consider it to be less environmentally damaging than the widely used bottom-trawling method. Scientists at the International Council for the Exploration of the Sea have so far found no evidence that the electric fishing methods in use have major negative impacts, but advise that more research is needed. The ban is not yet legally binding because the Parliament's fisheries committee must now work with the European Commission and EU member states to revise legislation.

PEOPLE

AIDS activist dies

Prominent AIDS researcher and activist Mathilde Krim died on 15 January at the age of 91. Krim (**pictured**) studied cancer and viral infections, focusing on the proteins the body makes to combat viruses. In 1983, she founded the AIDS Medical Foundation in New York City, a research and advocacy charity now called the Foundation for AIDS Research. Krim helped to raise public awareness of the AIDS epidemic in its early years, and campaigned for increased



funding for research into the condition, as well as for public-health programmes to reduce HIV transmission.

PUBLISHING

Preprint servers

Researchers can now share research articles written in Arabic and French, respectively, on two new preprint servers. The sites, Arabixiv and Frenxiv, will host manuscripts in many scientific disciplines. They were founded by Khaled Moustafa of the Paris-based National Conservatory of Arts and Crafts to address the scarcity of online scientific content in Arabic and French. The servers were built in partnership with the non-profit Center for Open Science in Charlottesville, Virginia.

Elsevier deals

After long negotiations, a Finnish university consortium has reached a deal with scientific publisher Elsevier

over access to paid journal content. The FinELib consortium had sought a nationwide journal-access agreement with the Dutch publisher after a row over increasing subscription prices. On 17 January, Elsevier said it had struck a three-year deal with FinELib that will allow 35 Finnish institutions access to about 1,850 journals on Elsevier's online database ScienceDirect. FinELib says the deal is valued at around €27 million (US\$33 million). In the previous week, Elsevier reached a similar deal with a consortium of 300 South Korean universities and libraries that had complained about price hikes. German institutions are still engaged in long-running negotiations over a nationwide licence.

RESEARCH

Shining remnants

The remains of the neutron-star merger that mesmerized astronomers last year continued to brighten until the end of the year, researchers reported last week (J. J. Ruan *et al. Astrophys. J. Lett.* **853**, 1; 2018; and R. Margutti *et al. Preprint at* <https://arxiv.org/abs/1801.03531>; 2018). Data from the Hubble Space Telescope, the Chandra X-ray Observatory and other telescopes suggest that a shockwave of matter

ejected in the collision — which took place 130 million years ago and was detected on 17 August through gravitational waves — is radiating with increasing intensity across the electromagnetic spectrum as it expands in the interstellar medium. Other observations from a European X-ray probe suggest that the brightness has peaked (P. D'Avanzo *et al. Preprint at* <https://arxiv.org/abs/1801.06164>; 2018). Researchers are awaiting Chandra's latest data to test those findings.

Monkeys cloned

Biologists in China have created the first primates cloned with a technique similar to that used to create Dolly the sheep (Z. Liu *et al. Cell* <http://dx.doi.org/10.1016/j.cell.2018.01.020>; 2018). Researchers hope to use the revised method to develop genetically identical primate populations to provide improved animal models of human diseases such as cancer. See page 387 for more.

Climate report

Last year was the third-warmest year on record, behind 2015 and 2016, according to an analysis released on 18 January by the US National Oceanic and Atmospheric Administration (NOAA). However, NASA, which used a different analysis, ranked 2017 as the second-warmest year on record, behind 2016. Both reports agree that record high temperatures around the world confirmed a long-term warming trend. According to NOAA, the average global temperature was 0.84°C above the twentieth-century mean. The NASA analysis used the reference period 1951–80 and found average temperatures to be 0.9°C higher than the global mean. Both analyses showed that the five warmest years on record have all taken place since 2010.

► **NATURE.COM**

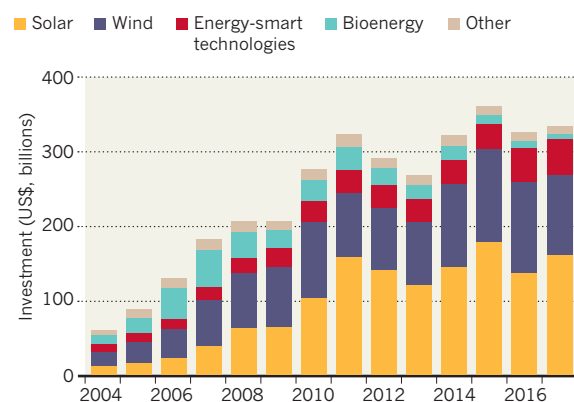
For daily news updates see:
www.nature.com/news

TREND WATCH

Global investment in clean energies totalled US\$333.5 billion last year, up by 3% from 2016. Solar energy attracted 48% of the total, notes a 16 January report. The growth was driven, in part, by a boom in installations of solar photovoltaic cells in China, which had a record year for clean-energy investment, spending \$132.6 billion. Spending in Britain dropped by 56% owing to policy changes, and by 26% in Germany. Global cumulative investment in clean energies amounts to \$2.5 trillion since 2010.

CASH FOR CLEAN ENERGY

Global investment in clean energy totalled US\$333.5 billion last year, the second-highest annual figure ever.

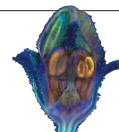


NEWS IN FOCUS

FUNDING German researchers expect more funding under new government **p.388**

METRICS China moves into first place in production of scientific articles **p.390**

ANTHROPOLOGY Questions emerge about oldest known hominin **p.391**



BOTANY Classical plant morphology is poised for a renaissance **p.396**

QIANG SUN/MU-MING POO/CAS



A baby long-tailed macaque, named Zhong Zhong, is one of the first cloned monkeys.

GENETICS

Monkeys cloned in China

Genetically identical animals promise improved models of human disease, but raise concerns about reproductive cloning of humans.

BY DAVID CYRANOSKI

Biologists in Shanghai, China, have created the first primates cloned with a technique similar to the one used to clone Dolly the sheep and nearly two dozen other species. The method has failed to produce live primates until now.

Researchers hope to use this revised technique to develop populations of genetically identical primates to provide improved animal models of human disorders, such as cancer. The technology, described in *Cell* on 24 January (Z. Liu *et al.* *Cell* <http://dx.doi.org/10.1016/j.cell.2018.01.020>; 2018), could

also be combined with gene-editing tools such as CRISPR–Cas9 to create genetically engineered primate-brain models of human disorders, including Parkinson's disease.

“This paper really marks the beginning of a new era for biomedical research,” says Xiong Zhi-Qi, a neuroscientist who studies brain disease at the Chinese Academy of Sciences Institute of Neuroscience (ION) in Shanghai. He was not involved in the cloning project.

But the achievement is also likely to raise some concerns among scientists and the public that the technique might be used to create cloned humans. “Technically, there is no barrier to human cloning,” says ION director

Mu-Ming Poo, who is a co-author of the study. But ION is interested only in making cloned non-human primates for research groups, says Poo: “We want to produce genetically identical monkeys. That is our only purpose.”

Primates have proved tricky to copy, despite many attempts using the standard cloning technique. In that method, the DNA of a donor cell is injected into an egg that has had its own genetic material removed.

ION researchers Sun Qiang and Liu Zhen combined several techniques developed by other groups to optimize the procedure. One trick was to undo chemical modifications in the DNA that occur when embryonic cells turn ▶

► into specialized cells. The researchers had more success with DNA from fetal cells, rather than cells from live offspring.

Using fetal cells, they created 109 cloned embryos, and implanted nearly three-quarters of them into 21 surrogate monkeys. This resulted in six pregnancies. Two long-tailed macaques (*Macaca fascicularis*) survived birth: Zhong Zhong, now eight weeks old, and Hua Hua, six weeks. Poo says that the pair seem healthy so far. The institute is now awaiting the birth of another six clones.

Cloning specialist Shoukhrat Mitalipov of the Oregon Health and Science University in Portland says that the Chinese team should be congratulated. “I know how hard it is,” says Mitalipov, who estimates he used more than 15,000 monkey eggs in cloning attempts in the 2000s. Although he was able to produce stem-cell lines from cloned human and monkey embryos, his team’s primate pregnancies never resulted in a live birth.

Cloned animals offer some significant advantages over non-clones as models for studying human disease. In experiments with non-cloned animals, it is difficult to know whether differences between the test and

control groups were caused by the treatment or genetic variation, says Terry Sejnowski, a computational neurobiologist at the Salk Institute for Biological Studies in La Jolla, California. “Working with cloned animals greatly reduces the variability of the genetic background, so fewer animals are needed,” he says.

PARKINSON’S STUDIES

Sejnowski also says that primate brains are the best model for studying human mental disorders and degenerative diseases. The ability to clone monkeys might revive primate studies, which have declined in most countries, says Poo. Parkinson’s disease experiments that currently use hundreds of monkeys could be done with just ten clones, he says.

Neuroscientist Chang Hung-Chun, also at ION, says that primate-cloning technology will soon be combined with gene-editing tools to study human genetic disorders in primate brains. Gene editing is already used on developing monkey embryos, but that leaves open the possibility that some cells are not edited, which then affects the results, says Chang.

With cloning, the donor cell can be edited before it is injected into the egg. Within a

year, Poo expects the birth of cloned monkeys whose cells have been genetically edited to model circadian-rhythm disorders and Parkinson’s disease.

Spurred by the promise of primate research, the city of Shanghai is planning major funding for an International Primate Research Center, expected to be formally announced in the next few months. The centre will produce clones for scientists around the globe. “This will be the CERN of primate neurobiology,” Poo says. There’s already high demand from pharmaceutical companies that want to use cloned monkeys to test drugs, he says.

Although most reproductive biologists are unlikely to consider using the technique to clone humans because of ethical objections, Mitalipov worries that it might be attempted in a private clinic.

China has guidelines that prohibit reproductive cloning, but no strict laws. It also has a weak record of enforcement of its rules on the use of stem cells for therapy. Some other countries — notably the United States — do not prohibit reproductive cloning at all. “Only regulation can stop it now,” says Poo. “Society has to pay more attention to this.” ■

GERMANY

Incoming government set to splurge on science

German spending may reach 3.5% of gross domestic product.

BY QUIRIN SCHIERMEIER

German politicians seem close to agreeing on a coalition government that would further boost federal funds for research — cementing the country’s status as one of the world’s biggest science spenders.

Political negotiations have been ongoing for four months since an inconclusive general election last September. In that election, Chancellor Angela Merkel’s centre-right Christian Democratic Union (CDU) gained the largest share of seats but no outright majority, and the Social Democratic party (SPD) — Merkel’s coalition partner in the last government — came second, and vowed to oppose the CDU rather than support it in government. After talks between the CDU and smaller parties broke down, the SPD voted on 21 January to seek to enter a coalition government again.

The parties have already set out the cornerstones of their coalition agreement in a paper leaked to the press on 12 January. These include injecting at least an extra €2 billion (US\$2.5 billion) of federal spending into

Germany’s science system over the coming years, in a bid to increase the country’s overall research spending from just under 3% of gross domestic expenditure to 3.5% by 2025. This would bring Germany into third place globally on the proportion spent on research and development, behind only Israel and South Korea. However, the German goal relies on contributions from the nation’s 16 state governments and industry to increase spending, as well as the federal government.

During Merkel’s 12-year chancellorship, federal science spending has almost doubled. Moreover, an agreement in 2005 between the federal government and state governments guaranteed annual budget increases of at least 3% to the country’s main science organizations — including the Max Planck Society, the Helmholtz Association of German Research Centres and the German Research Foundation (DFG), Germany’s main grant-giving agency for university research.

“All the indications are that research support remains a top government priority in many fields,” says Otmar Wiestler, president of the

Helmholtz Association in Berlin. “That’s very encouraging. Planning security is a prerequisite for us to be able to develop strategic research activities in key areas, such as mobility, climate change, energy supply, personal medicine and information technology.”

However, low public acceptance of genetic engineering in plants and the use of genetically modified organisms in agriculture remains a concern, says Jörg Hacker, president of the Leopoldina, Germany’s National Academy of Sciences in Halle. “Germany needs a bio-science agenda,” he says. “A technology-friendly society should be open to the potential of advances such as CRISPR–Cas technology.”

The coalition partners promised to improve funding opportunities for basic research into pressing societal challenges, including energy, health, mobility and security. Details have yet to be announced, but many scientists hope that the government might create a federal funding agency for blue-skies research.

The parties have also already set out plans to increase wind and solar energy capacity by about 10% by 2020. Germany currently meets about one-third of its electricity demand from wind, solar, hydro and biomass sources. However, it is currently expected to miss its goal of reducing carbon dioxide emissions by 40% relative to 1990 levels by 2020. Coalition partners have said that they will strive to produce at least 65% of Germany’s power generation with renewable energy sources by 2030 — twice current levels — and they have announced a legislative initiative to make Germany’s climate and energy targets legally binding. ■

POLITICS

Brief US shutdown ends

But science agencies face daunting possibility of another funding lapse next month, when temporary spending deal expires.

BY LAUREN MORELLO, SARA REARDON AND HEIDI LEDFORD

Scientists across the United States heaved a sigh of relief on 23 January, as the US government resumed operations after a three-day shutdown.

The impasse began on 20 January, after Congress let a temporary funding bill expire. The National Institutes of Health and the National Science Foundation prepared to stop processing grants, and many federal science agencies instructed ‘non-essential’ researchers to ready their labs and offices for indefinite closure. The previous government shutdown, in October 2013, lasted for 16 days — cutting short the US Antarctic Program’s field season, delaying some grant-funding cycles by several months and disrupting an untold number of carefully planned experiments.

This time, researchers were luckier: on 22 January, the Senate and House of Representatives approved a stopgap spending bill to cover government operations until 8 February, as lawmakers try to resolve major differences over immigration policy. But that quick fix will fund the government for less than three weeks, raising the possibility of another spending showdown in early February.

Jennifer Zeitzer, director of legislative relations at the Federation of American Societies for Experimental Biology in Bethesda, Maryland, says that she is “cautiously optimistic” about the progress that lawmakers have made in the past week towards a long-term budget agreement. “I’m going to withhold my panic for now,” she says.

Ideally, Zeitzer says, Congress would pass legislation by 8 February to raise limits on federal spending — clearing the way for a spending bill to cover the remainder of the 2018 fiscal year, which ends on 30 September. “I’m just hoping the experience of going through a shutdown was painful enough for everyone,” Zeitzer says.

In the meantime, researchers are waiting to see how the brief shutdown and continuing budget uncertainty might affect their work.

Peter Neff, a glaciologist at the University of Washington in Seattle, is part of a team that has conditional approval for an NSF grant to study trace gases in Antarctic ice. The researchers had hoped for the grant to start on 1 January, but it has been delayed by several weeks. Neff isn’t sure why — but he says that at a scientific meeting



Lawmakers in the US Congress agreed to a temporary budget for the government.

in December, an official with the NSF’s polar-programmes division said that it was operating under the assumption that it could face a 10% budget cut in the near future. Like other federal agencies, the NSF has been supported by a string of short-term spending measures since the 2018 budget year began in October.

For Neff’s team, any additional delay could make it difficult to plan field-season logistics with their international partners. “We’ve already put two years into this project, and we’re not going to have samples or data for another two years,” he says. “We don’t want that timeline to get any wider for any reason.”

The grant is also supposed to pay for 50% of Neff’s salary during his postdoctoral fellowship; until it comes through, the University of Washington is covering the full amount. “There are people who are in far more difficult situations than I am,” he says. “I can carry on and assume that everything will work itself out. But it’s just not an efficient way to operate.”

The shutdown’s end came just in time for Chad Hayes, a plant scientist at the US Department of Agriculture, to make a planned

trip this week to Mexico. There, his team intends to breed experimental sorghum at a winter nursery — the culmination of a year of planning.

Hayes expects to finish the field work by 8 February, but has to return to Mexico in March or April to harvest seeds from his sorghum plants and bring them back to the United States. If there is another shutdown then, the plants will go to waste. “When the plants say something’s ready, we have to be there,” Hayes says. “Plants know nothing about about weekends, holidays, or even government shutdowns!”

Others note that even planning for a shutdown creates major work for federal agencies, as they prepare to send employees home and shut down systems. Senior managers must think about which functions are crucial and how to justify continuing those if funding runs out, says Heather Howell, a former deputy director at the US Food and Drug Administration.

“Intense planning goes into every one of these possible shutdowns,” says Howell, now a consultant for NSF International in Washington DC. “It’s extremely costly to do all of that planning. Every time I sat in one of these meetings, I would wish that somebody would do an analysis of how much money is sitting around this table right now.” ■

METRICS

China declared largest source of research articles

Report also finds United States is still a science powerhouse despite increasing competition.

BY JEFF TOLLEFSON

For the first time, China has overtaken the United States in terms of the total number of science publications, according to statistics compiled by the US National Science Foundation (NSF).

The agency's report, released on 18 January, details the United States' increasing competition from China and other developing countries that are raising their investments in science and technology. But the report suggests that the United States remains a scientific powerhouse, pumping out high-profile research, attracting international students and translating science into valuable intellectual property.

"The US continues to be the global leader in science and technology, but the world is changing," says Maria Zuber, a geophysicist at the Massachusetts Institute of Technology in Cambridge. As other nations increase their output, the United States' relative share of global science activity is declining, says Zuber, who chairs the National Science Board, which oversees the NSF and produced the report. "We can't be asleep at the wheel."

The change (see 'Shifting landscape') is clear already in terms of the volume of publications: China published more than 426,000 studies in 2016, or 18.6% of the total documented in the NSF analysis of Elsevier's Scopus database. That compares with nearly 409,000 by the United States. India surpassed Japan, and the rest of the developing world continued its upward trend.

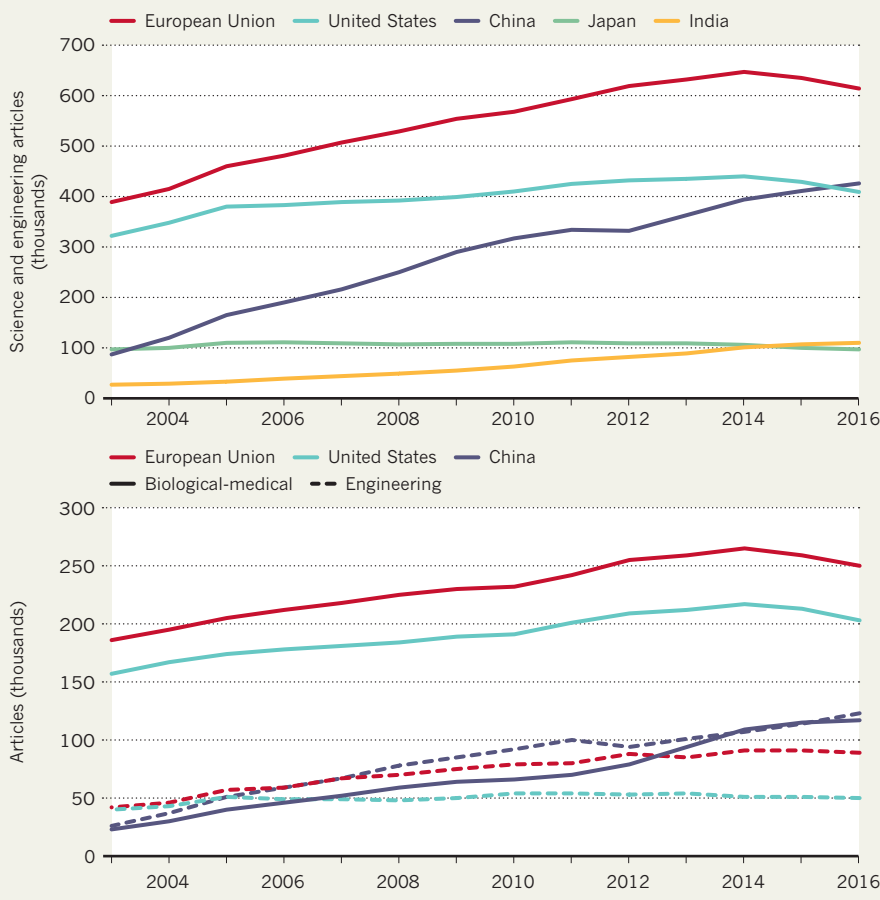
The NSF analysis divides the credit for a publication fractionally among its authors. By contrast, Scopus gives one full credit to each author; as a result, it still ranks the United States first in terms of the number of publications.

The United States ranked third, behind Sweden and Switzerland, when the NSF team examined where the most highly cited publications came from. The European Union came in fourth and China fifth. The United States still produces the most doctoral graduates in science and technology, and remains the primary destination for international students seeking advanced degrees — although its share of such students fell from 25% in 2000 to 19% in 2014, the report says.

The United States spent the most on research and development (R&D) — around US\$500 billion in 2015, or 26% of the global

SHIFTING LANDSCAPE

China now produces more scientific research articles per year than any other single nation, according to an analysis by the US National Science Foundation. The country outranks the United States in production of engineering articles, but lags behind on publications related to biomedical research.



SOURCE: NSF

total. China came in second, at roughly \$400 billion. But US spending remained flat in terms of its share of the country's economy, whereas China has increased its R&D spending, proportionally, in recent years.

The NSF analysis, the latest edition of the agency's biennial Science and Engineering Indicators, comes at a time of heightened concern about the state of US science. It should raise some alarms, says Mark Muro, a senior fellow with the Brookings Institution, a think tank in Washington DC. Trends in US science spending are heading in the wrong direction, he says, and the talent pool of researchers continues to be limited by under-representation of women

and minorities. Similarly, key industries such as semiconductor manufacturing have been hollowed out as businesses ship production work to other countries, Muro adds.

For the first time, the NSF included a section on technology transfer and innovation in its statistical analysis. Data suggest that the United States continues to lead the world when it comes to measures like patents, revenue from intellectual property and venture-capital funding for innovative technologies. "A nation's innovation capacity is one of the main drivers of productivity growth and so prosperity," Muro says. The new data provide "a useful reminder of why we care about these indicators in the first place". ■



A fossilized skeleton was discovered in Chad in 2001. Researchers have raised questions about its femur (long bone, centre right).

PALEOANTHROPOLOGY

Femur findings remain a secret

Fresh take on human ancestry struggles to be accepted.

BY EWEN CALLAWAY

When anthropologists meet in France at the end of January, one of the most provocative fossils in the study of human evolution will not feature on the agenda. The approximately 7-million-year-old femur¹ was examined more than a decade ago by scientists in the French city of Poitiers, but has yet to be thoroughly described in a published scientific paper.

The fossil may belong to the earliest known hominin, the group that includes humans and their extinct relatives. Few people have had access to it, but two scientists who analysed the bone briefly in 2004 have prepared a preliminary description of it. They had hoped to present their analysis at the meeting, which is organized by the Anthropological Society of Paris and takes place in Poitiers. But the proposal by Roberto Macchiarelli, a palaeoanthropologist at the University of Poitiers, and Aude

Bergeret, director of the Museum of Natural History Victor-Brun in Montauban, France, was rejected by the conference organizers.

"This specimen is really important. It's critical," says Macchiarelli, who has shared his unpublished report with *Nature's* news team. The femur probably belongs to a species called *Sahelanthropus tchadensis*, he says. The bone is important because it could settle whether the species is the earliest hominin yet found, as its discoverers have claimed after analysing the skull². "This is a fantastic occasion to finally tell people what we have, and what we know about this specimen."

The Anthropological Society of Paris told *Nature* that it had rejected 6 out of 65 abstracts. It said: "This work is conducted by an independent and impartial scientific committee, which is sovereign in its decision. Hence, any accusation about this would not be founded."

The *Sahelanthropus* femur was discovered early on the morning of 19 July 2001, beside a battered skull and other bones at a site in the Djurab Desert in northern Chad, says Alain ►



**MORE
ONLINE**

TOP NEWS



Indian start-up's chances for Moon-mission competition sink go.nature.com/2n3emnb

MORE NEWS

- Simple blood test detects eight different kinds of cancer go.nature.com/2n54gar
- 'Dark matter' DNA influences brain development go.nature.com/2rsujpw
- Brexit vote didn't spur quick academic exodus go.nature.com/2dxphgd

NATURE PODCAST



A mini all-terrain robot; 3D painting with light; and a new maze for rats nature.com/nature/podcast

► Beauvilain, a retired geographer who led the field team that made the discovery.

Michel Brunet, a palaeontologist at the University of Poitiers, who headed the Chadian expedition that discovered the *Sahelanthropus* remains, argues that the species is the earliest known representative of the hominin lineage.

His team described the skull — dubbed Toumai, which means ‘hope of life’ in the Chadian Daza language — in a 2002 *Nature* paper² that became a scientific blockbuster. A subsequent analysis of the skull and other fragments by Brunet and his team suggests that Toumai probably walked upright on two legs³. Brunet declined to comment on the analysis of the thigh bone or on Macchiarelli’s and Bergeret’s efforts to describe it at the Poitiers meeting. “Our studies are still in progress,” he wrote in an e-mail. “Nothing to say before publishing.”

Other researchers have questioned whether Toumai was indeed part of the lineage that led to humans, pointing to recently discovered fossils from Ethiopia and Kenya as better contenders for the earliest hominin. But Brunet’s team has stood by Toumai’s hominin status in response to the controversy⁴ and in a subsequent publication that described a lower jaw and teeth³.

Beauvilain says that the femur and other material remained in Chad until they were eventually shipped to Poitiers in 2003, where they were stored in a collection of animal-bone

fragments from the trip. In 2004, Bergeret, who was then a graduate student at the University of Poitiers, came across the blackened and badly damaged bone while analysing other bones in the collection. “I discovered the femur by chance,” she says.

EXCITING FIND

Brunet and other members of his team were back in Chad when Bergeret found the femur. So she asked Macchiarelli, who studies human evolution and who was then head of the department of geosciences at the University

“This is a fantastic occasion to finally tell people what we have, and what we know.”

of Poitiers, for help in analysing it. She says that she examined it closely for several days, comparing it to other hominin fossils. “I remember joking with another student, who told me, ‘You found Toumai’s femur!’,” Bergeret says. “I realized when I saw Roberto Macchiarelli that this joke was probably based on reality.”

In their short description of the femur, Macchiarelli and Bergeret contend that the bone differs greatly from that of a roughly 6-million-year old potential hominin found in Kenya in 2000 that is thought to have walked on two feet. Macchiarelli doubts that *Sahelanthropus*

is a hominin, but thinks a conclusion should be made only after more careful study of all its remains, including the femur.

The femur and other *Sahelanthropus* remains are crucial to determining the status of the species, because individual anatomical parts can often be misleading about evolutionary history, says Bernard Wood, a palaeoanthropologist at George Washington University in Washington DC. He says the fossil could belong to a now-extinct lineage of great ape.

A paper describing the femur is “long overdue”, says palaeoanthropologist Bill Jungers, at Stony Brook University in New York. “We don’t know why it’s been kept secret. Maybe it’s not even a hominin. Who the hell knows until someone can expose it.” ■

1. Lebatard, A.-E. et al. *Proc. Natl Acad. Sci. USA* **105**, 3226–3231 (2008).
2. Brunet, M. et al. *Nature* **418**, 145–151 (2002).
3. Zollikofer, C. P. E. et al. *Nature* **434**, 755–759 (2005).
4. Brunet, M. *Nature* **419**, 582 (2002).

CORRECTION

The Editorial ‘Vaccine boosters’ (*Nature* **553**, 259–250; 2018) said that the HIV-infected blood transfusions were given in the early 1990s. In fact, they were given in the 1980s.



CATACLYSM'S END

A popular theory about the early Solar System comes under fire.

BY ADAM MANN

Early in Earth's history, roughly half a billion years after the planet formed, all hell broke loose in the inner Solar System. A barrage of asteroids — some the size of Hong Kong — pummelled the globe intensely enough to melt large parts of its surface. This incendiary spree around 4 billion years ago vaporized most of Earth's water and perhaps even sterilized its exterior, killing off any life that might have started to emerge. Only after this storm of impacts passed did the planet become safe enough for hardy organisms to take firm root and eventually give rise to all later life.

That horrific episode, known as the Late Heavy Bombardment (LHB), has been an integral part of Earth's origin story for decades,

ever since geologists did a systematic study of samples brought back from the Moon by NASA Apollo missions. But now, the once-popular theory has come under attack, and mounting evidence is causing many researchers to abandon it. A growing community of planetary scientists thinks that things quietened down relatively quickly, with a steadily decreasing rain of asteroids that ended a few hundred million years after Earth and the Moon formed.

Settling the debate could have major ramifications for some of the biggest questions in geoscience: when did life emerge and what were conditions like on early Earth? But some researchers think that fresh samples will be needed to finally put this conundrum to rest.

They are looking with hope at the United States' recent pledge to send astronauts back to the Moon — although no timeline has yet been set. In the meantime, the community is grappling with the fact that a key chapter of Solar System history might be vanishing before their eyes.

"The Late Heavy Bombardment was seen as one of the great triumphs of the Apollo era," says geochemist Mark Harrison of the University of California, Los Angeles. "There's no question that something has happened in the past few years that has profoundly upset the apple cart."

The Solar System formed some 4.6 billion

An artist's impression of the early Earth, bombarded by Solar System debris.

NASA/GODDARD IMAGE LAB

years ago, after the centre of a massive cloud of gas and dust collapsed into a dense sphere that became our Sun. Pebbles in a dusty disk orbiting the star continuously collided and sometimes stuck together. After tens of millions of years, these agglomerations had built up into planetesimals — the beginnings of the planets. Other rocky fragments remained, crashing into their larger kin and leaving deep craters. Over time, the Solar System thinned out, leaving something like the configuration we see today.

Most of the evidence of this violent history has been erased on Earth by the churning of tectonic plates. But the scarred surface of the Moon, long inert, retains a lengthy record of impacts. Some of that record — roughly 382 kilograms of lunar rock and soil — was collected by Apollo astronauts and carried back to scientists eager to see what the samples might reveal about the Moon's history. In 1973, the year after the last Apollo landing, a group at Sheffield University, UK, reported a curious pattern in samples from four separate Apollo missions as well as a Soviet Luna mission. Radiometric dating of each one returned the same age: 3.95 billion years¹. A team at the California Institute of Technology (Caltech) in Pasadena corroborated the findings the same year².

CURIOUS CHRONOLOGY

The confluence of ages suggested that a flurry of objects struck the Moon in a narrow 50-million-year window, leaving behind countless impact craters — including as many as a dozen of the Texas-sized basins that scar the surface. Because it seemed to represent a final surge of pandemonium after the Solar System's chaotic genesis, the Caltech team named the event the terminal lunar cataclysm, although it later became more popularly known as the LHB.

The idea was immediately divisive, in large part because of ambiguity in the rock dating. This was done primarily by measuring the rocks' ratio of argon-40 atoms to radioactive potassium-40. ⁴⁰K decays into ⁴⁰Ar with a half-life of 1.25 billion years. At high temperatures, that ⁴⁰Ar can leak out of minerals. That makes the ratio of these two isotopes a kind of clock: the more time that has elapsed since a rock was hot, the more ⁴⁰Ar should be present. But making sense of the argon and potassium

concentrations can be difficult because the same ratio could have been caused by a concentrated barrage that heated the rocks and released ⁴⁰Ar some 3.95 billion years ago, or by a long, dwindling asteroid torrent that released it in fits and starts before fizzling out at about the same time.

The first really new data arrived in 2000. Planetary scientist David Kring, cosmochemist Timothy Swindle and planetary scientist Barbara Cohen, all then at the University of Arizona in Tucson, collected lunar meteorites that had fallen to Earth after being blasted from the Moon's surface by asteroid strikes. They hoped such rocks would provide a more random sample of the Moon's crust than those from Apollo, which represent at most 4% of the lunar surface. But when the results came back, they showed a curious, and familiar, pattern.

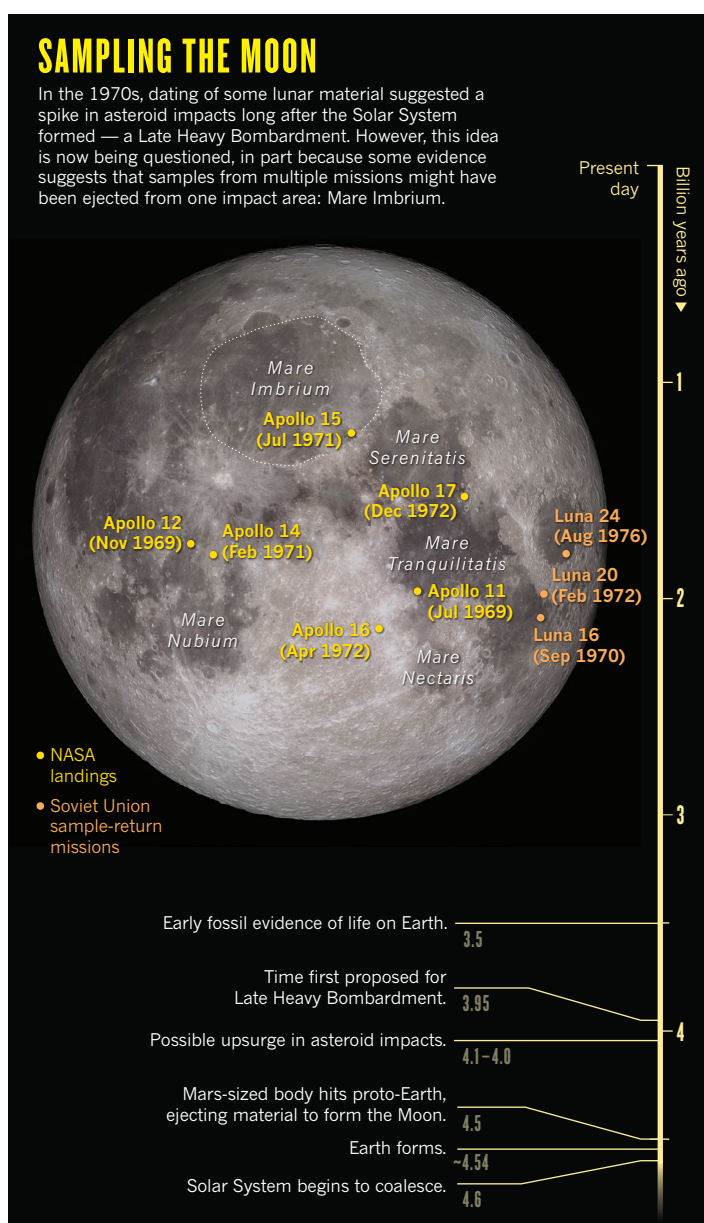
"Frankly, I thought we'd measure a bunch of these and have ages running back to 4.3 and 4.4 [billion years] and prove once and for all that

remote reservoirs and kicked asteroids out of the main belt.

The Nice model offered huge support for the LHB. "I think this helped cement this idea," says physicist Nicolle Zellner of Albion College in Michigan. Geologist Marc Norman of the Australian National University in Canberra agrees. "That was the next real turning point," he says.

CATACLYSMIC CONFUSION

Yet just when the idea of the LHB finally seemed unimpeachable, holes began to appear. Apollo data and 'crater counting', which estimates the order in which craters were laid down on the basis of how they overlap, had indicated that three of the largest crater basins on the Moon's near side — Imbrium, Nectaris and Serenitatis — might all be about 3.95 billion years old (see 'Sampling the Moon'). But high-resolution maps from NASA's Lunar Reconnaissance Orbiter, which



started circling the Moon in 2009, spotted rays of debris extending from Imbrium⁵. This suggested that the impact that formed the crater might have knocked rocks into nearby Serenitatis, contaminating the Apollo samples picked up there. In 2010, a reanalysis of rocks thought to have been ejected from Nectaris indicated that they were also chemically and geologically similar to Imbrium material⁶. “We started realizing that maybe we were sampling Imbrium over and over,” says Zellner.

The data from lunar meteorites didn’t necessarily help. Although none of the samples seemed to be older than 4 billion years, some were billions of years younger than that³, with no obvious spike around 3.95 billion years. And the Apollo samples held other surprises. Since 2012, detailed study⁷ of microscopic regions in the rocks has turned up ages of as much as 4.2 billion years, much older than any seen before, suggesting that there had been significant impacts earlier than the proposed spike.

Prodded in part by these revelations, some researchers proposed⁸ a longer-lasting LHB that began around 4.1 billion or 4.2 billion years ago. But that idea had one major strike against it: some of the most ancient crystals on Earth, from the Jack Hills range in Australia, suggest⁹ that the planet was a fairly clement place then, with relatively low temperatures and ample water.

HOT TOPIC

Others are still scrutinizing the original Apollo evidence. To determine the samples’ ages, researchers heated the rocks to release argon, slowly ramping up the temperature. But as far back as 1991, Harrison had pointed out that the process won’t work well for rocks containing multiple minerals. Different minerals will release their argon at different temperatures. A sample heated to 400 °C might provide an age of 2 billion years; to 500 °C, an age of 2.5 billion. Researchers have tried to extrapolate from this behaviour, but Harrison says the complex patterns often lead them to pick essentially arbitrary ages. “This is quackery,” he says. “There’s no physical basis for it.”

Swindle says the argon heating situation is not necessarily as bad as Harrison makes it out to be; Apollo samples can be found whose ages don’t change significantly with temperature, and their dates — whether they refer to one or multiple impacts — still cluster around 3.95 billion years. Cohen says that other chronometers, such as those using radioactive isotopes of rubidium and uranium, corroborate the argon ages (although Harrison counters that the dates can differ by as much as 600 million years).

Such back and forth underscores how difficult it can be to tease small clues out of extremely ancient rocks. “Sherlock Holmes was good at resolving mysteries that happened last year,” says David Nesvorný, a planetary scientist at the Southwest Research Institute in Boulder, Colorado. “This all happened 4 billion years ago.”

Meanwhile, the Nice model has proved less helpful to the idea of an LHB than it once seemed. More-advanced simulations of the early Solar System’s gravitational interactions indicate that the planetary reshuffling probably happened shortly after formation, not with a delay of hundreds of millions of years¹⁰. Nesvorný likens delaying the reshuffling — and so keeping the Solar System hovering on the edge of instability — to trying to balance a pencil on its tip. “It’s really hard to put the pencil there

“PEOPLE SEE WHAT THEY WANT TO SEE AND DISREGARD THE REST.”

in such a way that it falls in an hour,” he says.

One of the original architects of the Nice model, astronomer Alessandro Morbidelli of the Cote d’Azur Observatory in Nice, admits that the first versions took fine-tuning to get the reshuffling to occur so late. He no longer believes in the LHB, and sees many others in the field trading in the idea of a sudden asteroid deluge for that of a long, declining tail of bombardment. “My prediction is people will abandon the cataclysm,” he says.

Even those who remain tied to the LHB have had to modify their ideas. Planetary scientist William Bottke of the Southwest Research Institute agrees that there is no longer much support for a single, short spike. He says the best reading of the evidence, including samples from ancient Earth and radiometric dates in meteorite rocks, is a more drawn-out surge of bombardment that began around 4.1 billion or 4 billion years ago, with a relative lull before that, consistent with the existence of surface water in that period.

Astronomer William Hartmann, a visiting scientist at the International Space Science Institute in Bern, thinks the current situation proves that the idea of a cataclysm was never particularly robust. Various research communities “kind of had the impression that the other community had really solved this,” he says. “A paradigm structure was built up from supporting evidence, none of which was actually conclusive in itself.”

If an LHB did not happen, that could make it easier to explain how life emerged. Evidence of microbial life has been found in rocks that are around 3.5 billion years old. But those fossils seem quite complex, suggesting that they had been evolving from earlier forms for at least a few hundred million years, during the originally hypothesized time of the LHB. Without the cataclysm, such an ancient genesis might make more sense. Then again, some evidence

suggests that the microbes at the base of the tree of life were hyperthermophiles — that is, organisms that thrived in extreme heat. The intense conditions created by a rain of asteroids could have resulted in a number of pockets where life might have emerged.

So far, efforts to clinch the LHB debate with evidence from other likely victims — Mercury, Venus, Mars and objects from the asteroid belt — have proved inconclusive. Each camp accuses the other of cherry-picking favourable data and not looking at the total picture. “It’s a Rorschach test,” says Norman. “People see what they want to see and disregard the rest.”

The only thing that researchers say will substantially move the needle is new samples from the Moon. Kring, now at the Lunar and Planetary Institute in Houston, Texas, has developed some concepts for sample-return missions, including one that would see astronauts collecting rocks from the South Pole–Aitken basin, the largest and oldest impact crater on the Moon. However, the next human mission to the Moon is still a long way off. The first new lunar rocks to be carried back to Earth may come from China’s Chang’e-5, a robotic mission currently planned for 2019. It aims to collect samples from the volcanic Mons Rümker formation, an area younger than those explored by Apollo astronauts.

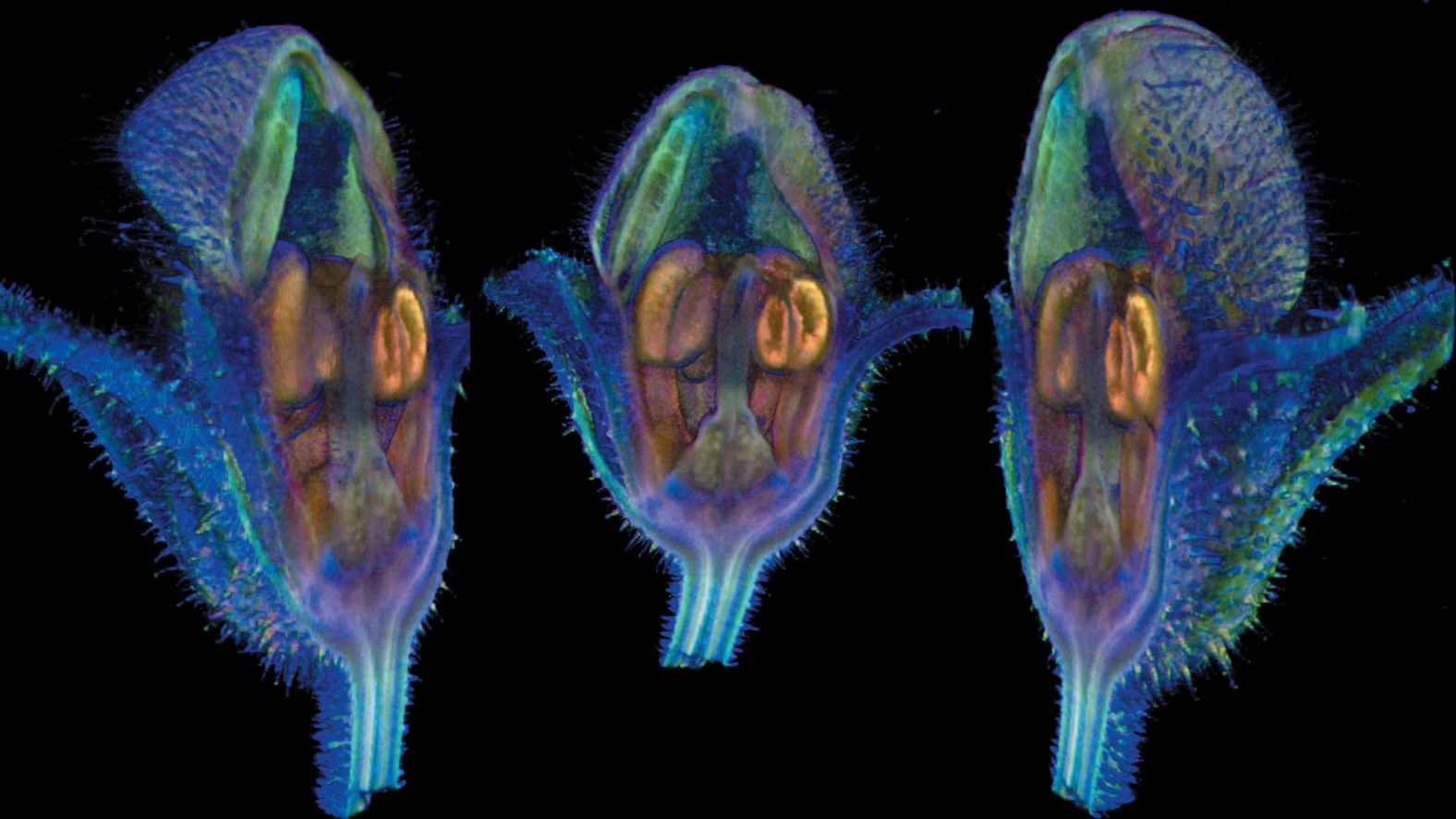
Although no single exploration effort is likely to end the dispute, researchers’ improved understanding of the Moon and how to determine the ages of samples should provide greater confidence in the results.

However things eventually shake out, the new evidence will shift careers and rewrite textbooks. Yet, perhaps because of the long-lived nature of this debate, those trying to make sense of the LHB remain flexible, sceptical and surprisingly lighthearted.

“We are close friends and therefore we disagree all the time and then go drink a beer together,” says Bottke. “One should carry models lightly and be prepared to drop them if something better comes along, because it happens all the time.” ■

Adam Mann is a freelance journalist based in Oakland, California.

1. Turner, G., Cadogan, P. H. & Yonge, C. J. *Proc. Lunar Sci. Conf.* **4**, 1889–1914 (1973).
2. Tera, F., Papanastassiou, D. A. & Wasserburg, G. J. *Abstr. Lunar Planet. Sci. Conf.* **4**, 723–725 (1973).
3. Cohen, B. A., Swindle, T. D. & Kring, D. A. *Science* **290**, 1754–1756 (2000).
4. Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. *Nature* **435**, 466–469 (2005).
5. Spudis, P. et al. *J. Geophys. Res. Planets* **116**, E00H03 (2011).
6. Norman, M. D., Duncan, R. A. & Huard, J. J. *Geochim. Cosmochim. Acta* **74**, 763–783 (2010).
7. Norman, M. D. & Nemchin, A. A. Early Solar System Impact Bombardment II Conf., Lunar and Planetary Institute (Houston), Abstr. 4013 (2012); available at go.nature.com/2mbdwvc
8. Morbidelli, A., Marchi, S., Bottke, W. F. & Kring, D. A. *Earth Planet. Sci. Lett.* **355–356**, 144–151 (2012).
9. Watson, E. B. & Harrison, T. M. *Science* **308**, 841–844 (2005).
10. Deienno, R., Morbidelli, A., Gomes, R. S. & Nesvorný, D. *Astron. J.* **153**, 153 (2017).



BOTANICAL RENAISSANCE

ADVANCES IN GENOMICS AND IMAGING
ARE REVIVING A FADING DISCIPLINE.

BY HEIDI LEDFORD

When Elizabeth Kellogg finished her PhD in 1983, she feared that her skills were already obsolete. Kellogg studied plant morphology and systematics: scrutinizing the dazzling variety of plants' physical forms to tease out how different species are related. But most of her colleagues had already pivoted to a new approach: molecular biology. "Every job suddenly required molecular techniques," she says. "It was like I had learned how to make illuminated manuscripts, and then somebody invented the printing press."

Kellogg had graduated near the start of a revolution in plant biology. Over the next few decades, as researchers adopted molecular tools and DNA sequencing, detailed analyses of plants' physical traits fell out of fashion. And because many geneticists worked with only a few key organisms, such as the thale cress *Arabidopsis thaliana*, they didn't need expertise in comparing and contrasting different plant species. At universities, botany departments folded and molecular-biology

3D imaging offers new views of antirrhinum (snapdragon) buds.

departments swelled. Kellogg, now at the Donald Danforth Plant Science Center in St Louis, Missouri, adapted: she embraced genomics, and combined it with her morphology skills to trace the evolution of key traits in the wild relatives of food crops.

But lately, Kellogg has noticed a resurgence of interest in the old ways. Advances in imaging technology — allowing researchers to peer inside plant structures in 3D — mean that biologists are seeking expertise in plant physiology and morphology again. And improvements in gene editing and sequencing have liberated geneticists to tinker with DNA in a wider range of flora, giving them a renewed appetite to understand plant diversity.

Plant biologists hope that, by combining new approaches to botany with data from genomics and imaging labs, they can provide better answers to questions that biologists have asked for more than 100 years: how genes and the environment shape the rich diversity of plants' physical forms. "People are starting to look beyond their own system into plants as a whole," says Kellogg. Plant morphology was once a science of form for its own sake, she says, but now, it is being pressed into service to understand how plant traits connect to gene activity across disparate species. "It's coming back — just under different guises."

BOTANY 2.0

Plant morphologists trace their roots back to the eighteenth-century German philosopher and poet Johann Wolfgang von Goethe, who took in the breadth of plant diversity and embarked on a search for an archetypal plant from which all forms could be derived.

That romantic idea went unfulfilled, but scientists continued his approach of comparing plant structures and functions to learn more about how they evolved and developed. The evolution of flowering plants would later trouble Charles Darwin, who famously called the rapid expansion of such a vast range of flower shapes, colours and pollination strategies an "abominable mystery".

Although the genomics era led many plant biologists away from morphology, the latest generation of technological advances is steering them back towards the questions that occupied Goethe and Darwin.

Prominent among these are computed tomography (CT) scanners, which can create 3D reconstructions of internal plant structures without destroying tissue. At the University of Vienna, for instance, plant morphologist Yannick Staedler has used CT scanners to analyse the secrets of a deceptive group of European orchids. Whereas many orchids reward insect pollinators with nectar, others imitate a mating partner or a nectar-rich flower but provide no reward. Biologists back to the time of Darwin have wondered how these 'deceptive orchids' thrive, because an insect is unlikely to visit them more than once. Staedler's studies suggest that such orchids might produce more ovules — the part of the ovary that becomes the seed — potentially to compensate for reduced pollination rates¹.

Erika Edwards, a plant morphologist at Yale University in New Haven, Connecticut, is using CT scanners to analyse how the shapes of leaves might be influenced by their early development inside the constrained space of a bud. Botanists have noted for a century that more-serrated, toothed leaves are found in northern, cold regions, whereas smoother leaves are seen in wet tropical forests — but it's still not clear why. Edwards hopes to unravel the connection.

Some researchers are combining 3D imaging and molecular tools. At the John Innes Centre in Norwich, UK, Enrico Coen's flower development laboratory uses a technique called optical projection tomography to capture 3D images of plants as they grow. It can also image

insect pollinators caught rummaging inside flowers or trapped inside a carnivorous plant. Simultaneously, the group is monitoring gene activity in the plants, by tagging key proteins with fluorescent markers. By combining classical morphology studies with 3D imaging and insights from developmental biology, the group hopes to learn more about the mechanisms that generate plant forms, Coen says. In one study², for example, he and his collaborators monitored barley-flower development, and explained why that process goes awry in a mutant of barley that was first discovered in the 1830s in Nepal.

Other new imaging techniques are aimed squarely at improving crop breeding. In a field in Jülich, Germany, drones and mini blimps mounted with thermal-imaging cameras fly over plants, while unmanned vehicles called FieldCops carry sensors as they patrol the ground. The effort, at the Jülich Plant Phenotyping Centre, is part of a growing movement to rapidly collect data about plant traits. Initially, these included a limited range of characteristics, such as growth rates or the number of seeds produced. But drones and robots have been fitted with increasingly sophisticated sensors, notes Dirk Inzé, a plant molecular biologist at Ghent University in Belgium. Some are now able to collect data about plant architecture, such as branching and leaf shape, using laser scanners and depth sensors. Similar scanners have been used in lab-grown plants to analyse the rhythmic growth of leaves, and to link that growth to a particular protein complex³.

FROM GENOMES TO PATTERNS

Molecular labs might also feel a pull back towards botany because, as in other areas of genomics, reading DNA has become so cheap that merely sequencing a plant species is no longer an end unto itself. The first published plant genome — that of *A. thaliana* — appeared in 2000, and more than 250 plant species have been sequenced since. Now, says William Friedman, director at the Arnold Arboretum of Harvard University in Boston, Massachusetts, "people want to ask how genomes explain evolution and pattern".

In 2017, for example, the publication announcing the genome of the orchid *Apostasia shenzhenica* included an analysis of genes that are likely to be responsible for unique aspects of orchid morphology. This includes the labellum, a part of the orchid flower that attracts insects and serves as a landing pad⁴.

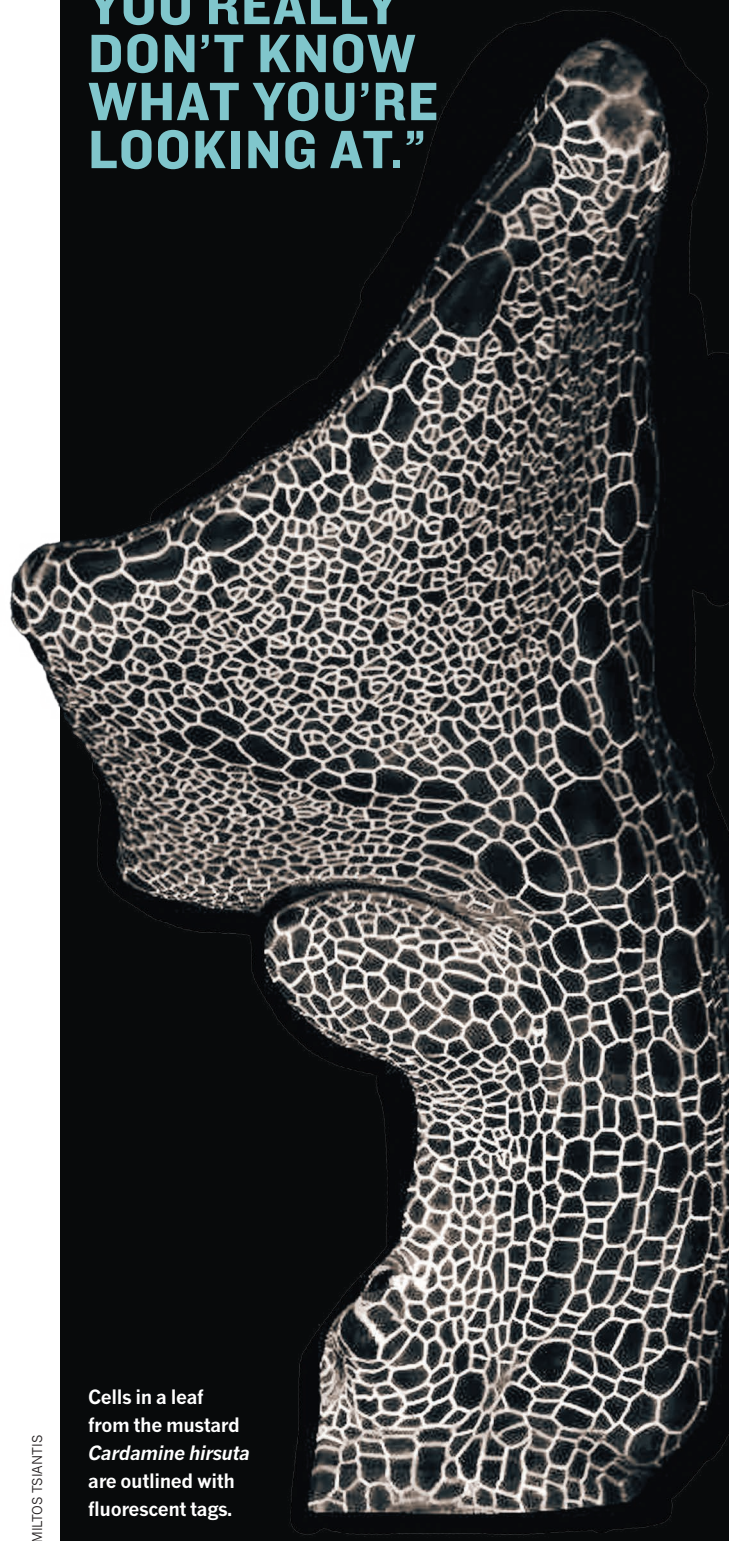
"It's possible now to understand the paths through which genetic changes influence form," says Miltos Tsiatis of the Max Planck Institute for Plant Breeding Research in Cologne, Germany. In 2014, his lab used genetics and time-lapse imaging to work out how a particular gene affects leaf shape by restraining cell growth at the leaf's edge in the mustard species *Cardamine hirsuta*⁵. Whereas *C. hirsuta*'s leaves grow as a series of leaflets around a stem, loss of this gene led to the simple oval leaves found in *A. thaliana*.

Plant morphologist Dan Chitwood, now at Michigan State University in East Lansing, harnessed sequencing power to look at gene expression in *Caulerpa taxifolia* — a seaweed that forms complex structures, including a stem and fern-like fronds, from a single, super-sized cell⁶. Some biologists have argued that the amount and rate of cell division is what shapes plant morphology. But Chitwood's study showed that gene expression in the unicellular seaweed varies in ways that echo gene expression in similar structures in multicellular plants — suggesting that the dividing cell needn't always dictate morphology.

Improved molecular tools have now made it possible to tweak DNA in plants that were previously too difficult to work with. The genome-editing tool CRISPR–Cas9 has enabled researchers to tinker with particular genes in a wide range of plants. Researchers have used it to turn purple morning glories white⁷, for instance, and to

**"IT'S
POSSIBLE
NOW TO
UNDERSTAND
THE PATHS
THROUGH
WHICH
GENETIC
CHANGES
INFLUENCE
FORM."**

“IF YOU’RE NOT REALLY CAPABLE OF DIAGNOSING MORPHOLOGY OR ANATOMY, YOU REALLY DON’T KNOW WHAT YOU’RE LOOKING AT.”



Cells in a leaf from the mustard *Cardamine hirsuta* are outlined with fluorescent tags.

MILTOS TSIAKIS

alter genes that are involved in building cell walls in orchids⁸.

But geneticists need to brush up on their botany skills to understand the full implications of these experiments, says Karl Niklas, who studies plant evolution at Cornell University in Ithaca, New York. Researchers often knock genes out to determine how they affect a plant's form or function. “If you're not really capable of diagnosing the morphology or the anatomy, you really don't know what you're looking at,” Niklas says.

He recalls a time when a student came to him with a mutant form of maize (corn) to show how the xylem — the collection of tubes that carry water and nutrients from the roots to the rest of the plant — was deformed. But the student was actually looking at normal phloem, a different network of vessels with a distinct structure that distributes nutrients formed in the leaves. “You know, it just makes your teeth hurt,” he says.

Researchers also lose out when they do not take the time to consider the diversity of plant forms in nature, says Chelsea Specht, a plant biologist also at Cornell University. She has seen cases in which scientists have failed to realize that their genetic mutants — for instance, *Arabidopsis* mutants with altered branching patterns — are recapitulating naturally occurring plant forms found in other lineages. When this happens, she says, researchers miss opportunities to put traits into an evolutionary context.

BOTANY BOOTCAMPS

The prospect of fading expertise so worried Friedman that, in 2013, he and his wife, plant morphologist Pamela Diggle of the University of Connecticut in Storrs, launched an intensive botany bootcamp for biologists. “It's been one of my missions as an academic to keep that knowledge going,” says Diggle. “It's important to keep this information alive in the community.”

The programme was first funded by the US National Science Foundation, and the New Phytologist Trust, a plant-science non-profit organization in Lancaster, UK, plans to pick up the bill from this year. It accepts about a dozen scientists each year, some from laboratories that typically focus on molecular biology and genomics. The course routinely has about six times as many applicants as positions, says Friedman.

Jamie Kostyun, an evolutionary geneticist, took the course in 2013 to gain the skills she needed to explore the floral traits of the genus *Jaltomata*. These species are kin to kitchen staples such as tomatoes and potatoes, but they boast a remarkable and recently evolved diversity of flowers. Some are flat, others tubular; some reward pollinators with sticky orange nectar, others ooze a blood-red sweet treat.

“They have crazy floral variation that nobody has looked at before,” Kostyun says. “I wanted to understand where that diversity came from.” She has used her plant-morphology training to detail the development of flowers in five *Jaltomata* species in her PhD thesis. Now, as a postdoc at the University of Vermont in Burlington, she is studying the panoply of nectar compositions and genetically analysing the vast array of flower shapes.

Friedman hopes that others will follow in Kostyun's footsteps, uniting these approaches with classical comparative techniques and generating insights into questions that have dogged researchers for decades. “What did the first flowers look like? You could probably open a book from 1900 and still ask the same questions that people were asking about basic plant structure,” he says. “We know more now, but we don't necessarily know the answers.” ■

Heidi Ledford is a senior reporter for Nature in London.

1. Staedler, Y. M. *et al. J. Exp. Bot.* <http://dx.doi.org/10.1093/jxb/erx405> (2017).
2. Richardson, A., Rebocho, A. B. & Coen, E. *Plant Cell* **28**, 2079–2096 (2016).
3. Dornbusch, T., Michaud, O., Xenarios, I. & Fankhauser, C. *Plant Cell* **26**, 3911–3921 (2014).
4. Zhang, G.-Q. *et al. Nature* **549**, 379–383 (2017).
5. Vlad, D. *et al. Science* **343**, 780–783 (2014).
6. Ranjan, A., Townsley, B. T., Ichihashi, Y., Sinha, N. R. & Chitwood, D. H. *PLoS Genet.* **11**, e1004900 (2015).
7. Watanabe, K. *et al. Sci. Rep.* **7**, 10028 (2017).
8. Kui, L. *et al. Front. Plant Sci.* **7**, 2036 (2017).

COMMENT

TECHNOLOGY From training to therapy — applications of virtual reality surveyed **p.402**

CULTURE Biography of YouTube maps a parallel universe of viral video **p.403**

PLASTICS China's ban on imported waste could boost sustainability **p.405**



OBITUARY Calestous Juma, African science champion, remembered **p.406**

ILLUSTRATION BY DAVID PARKINS



Repeating experiments is not enough

Verifying results requires disparate lines of evidence — a technique called triangulation. **Marcus R. Munafò** and **George Davey Smith** explain.

Several studies across many fields estimate that only around 40% of published findings can be replicated reliably. Various funders and communities are promoting ways for independent teams to routinely replicate the findings of others.

These efforts are laudable, but insufficient. If a study is skewed and replications

recapitulate that approach, findings will be consistently incorrect or biased. Consider a commonly used assay in which the production of a fluorescent protein is used to monitor cell activity. If the compounds used to manipulate cell activity are also fluorescent, as has happened¹, reliably repeatable results will not yield robust conclusions.

We have both spent much of our careers advocating ways to increase scientific certainty. One of us (M.R.M.) participated in work by UK funding agencies to develop strategies for reproducible science, and helped to craft a manifesto for reproducibility².

But replication alone will get us only so far. In some cases, routine replication might ►

► actually make matters worse. Consistent findings could take on the status of confirmed truths, when they actually reflect failings in study design, methods or analytical tools.

We believe that an essential protection against flawed ideas is triangulation³. This is the strategic use of multiple approaches to address one question. Each approach has its own unrelated assumptions, strengths and weaknesses. Results that agree across different methodologies are less likely to be artefacts.

Isn't this how science is meant to operate? Perhaps so, but scientists in today's hyper-competitive environment often lose sight of the need to pursue distinct strands of evidence.

The problem was aptly described in May 2017, when cancer researcher William Kaelin lamented that the goal of the scientific paper had shifted from testing narrow conclusions in multiple ways to making a broadening series of assertions, each based on limited evidence⁴. Consequently, he said, "papers are increasingly like grand mansions of straw, rather than sturdy houses of brick".

The scientific community should address this lack of depth strategically and establish practices that facilitate triangulation. Specifically, we advocate a system to support multidisciplinary teams, each created around a common question (see "Triangulation"). This, we believe, would result in robust insights — mansions of stone.

SPECIOUS ROBUSTNESS

We rarely see projects that aim to prove a point from multiple views. Psychology, epidemiology and the clinical sciences are all geared towards producing statistically significant, definitive studies centred on an endpoint that supports a hypothesis. In parts of the biological sciences, a manuscript's acceptance often depends on a 'capstone' study showing animal efficacy, so pursuing

TRIANGULATION

A checklist.

- The different approaches address the same underlying question.
- The key sources of bias for each approach are explicitly acknowledged.
- For each approach, the expected directions of all key sources of potential bias are made explicit, where feasible.
- Ideally, some of the approaches being compared will have potential biases that are in opposite directions.
- Ideally, results from more than two approaches — which have different and unrelated key sources of potential biases — are compared. [Source: ref. 3](#)



BSIP/UG/GETTY

It took many lines of evidence to show that maternal smoking results in babies with low birth weights.

that single experiment becomes more important than carefully probing an idea from all directions. Moreover, these studies are often presented as having implications for human health without including any tests in humans.

Although many studies in the basic sciences include some element of triangulation, they rarely do enough of it.

In our field of epidemiology, there are countless examples of spurious, persistent findings. Large observational studies frequently produce precise conclusions that are precisely wrong. A correlation between X and Y might be real in that it genuinely describes an observed association between variables, but is one that does not reflect cause and effect. No amount of replication or statistical adjustment can resolve this, and one of us (G.D.S.) has devoted more than two decades to developing methods that support stronger causal inference in observational epidemiology, drawing on disciplines from the basic sciences to economics.

An illuminating example is the oft-observed J-shaped curves that chart correlation between a condition and health outcome⁵.

For instance, multiple studies show that people who consume low levels of alcohol are healthier than heavy drinkers and teetotalers, leading several researchers to conclude that moderate alcohol consumption promotes health. But other factors, such as unhealthy people being advised to give up drinking, would explain the same shape. Similarly, repeated observations that being slightly overweight is associated with the highest life expectancy might be explained by illness (including processes leading up to the manifestation of a disease, which itself can result in reduced weight); by physicians treating overweight individuals more aggressively; and by other favourable characteristics

of overweight individuals, such as lower smoking rates.

How can one tell that a consistently observed relationship between a behaviour and a health outcome is causal? One example in which triangulation has helped is in establishing that smoking during pregnancy results in babies with lower birth weights⁶. That is different from the simple observation that women who smoke are more likely to have babies who weigh less. Smokers tend to have other characteristics that are also associated with low birth weight, such as low income, less education or more drug use.

Triangulation means explicitly choosing analytical approaches that depend on different assumptions. For example, if a woman's partner smokes during her pregnancy, many of the same confounders apply as in maternal smoking, but the association with lower birth weight is much weaker. Birth weight can also be analysed according to levels of cigarette taxation across US states, which reduces the effects of confounders. And analyses can compare the birth weights of siblings whose mother smoked during one pregnancy but not another.

Mendelian randomization is a technique developed specifically to probe causal relationships. In cohorts grouped according to whether or not people carry a genetic variant associated with greater cigarette consumption in those who smoke, mothers who smoke and carry the variant tended to have babies who weighed less; non-smokers with the same variant did not. Taken together, these studies make it clear that maternal smoking affects birth weight directly⁶.

REPLICATION FIXATION

Replication has received considerable attention; triangulation has not. Maybe one reason replication has captured so much

interest is the often-repeated idea that falsification is at the heart of the scientific enterprise. This idea was popularized by Karl Popper's 1950s maxim that theories can never be proved, only falsified. Yet few experiments, including replication attempts, are explicitly set up to falsify a theory. In fact, we worry that an overemphasis on repeating experiments could provide an unfounded sense of certainty about findings that rely on a single approach.

Moreover, philosophers of science have moved on since Popper. Better descriptions of how scientists actually work include what epistemologist Peter Lipton called in 1991 "inference to the best explanation", or the search for the "loveliest" explanation⁷. This draws on older ideas that championed abductive over deductive reasoning — looking for likely explanations rather than deriving explanations from first principles. This spirit is also captured in the idea of consilience put forward by polymath William Whewell in the mid-nineteenth century and popularized in the 1990s by naturalist E. O. Wilson. This posits that strong theories emerge from the synthesis of multiple lines of evidence, as when Charles Darwin proposed evolution by natural selection.

Unlike consilience, triangulation suggests the deliberate use of different methods. It is the approach to inference that aligns most closely with how many philosophers feel scientists come to understand reality. But most scientists would be hard-pressed to describe it. Researchers typically receive extensive training in experimental methods and design, but little in approaches to causal inference. They are left with no framework to guide scientific pursuit.

CREDIT SHIFT

Triangulation usually requires input from multiple methodologies or disciplines. An elegant historical example is continental drift. In the early 1900s, geophysicist Alfred Wegener noticed that the shape of the west coast of Africa seems to fit that of the east coast of South America. He sought evidence to support the continental-drift theory from a wide range of sources, such as palaeontology (fossils from the same period appeared on both continents) and geology (glacier markings indicated that the continents were once close). In today's environment, scientists would need to contribute to multi-disciplinary projects, with studies providing distinct lines of evidence.

Encouraging such an approach will require fundamental changes to the way in which credit is attributed and to how peer review is conducted. In the current system, few authorship positions count much towards credit — in biomedical science, say, it typically falls just to the corresponding

and other starred authors, as well as to first authors.

To support triangulation, we recommend a shift to a contributorship model, similar to the credits that roll at the end of a film — a long list of individuals with their contributions described fully and specifically⁸. This will require academics to potentially forgo 'senior authorship' positions. It would also make it easier for early-career researchers to specify their unique contribution to a paper when applying for promotion or another position.

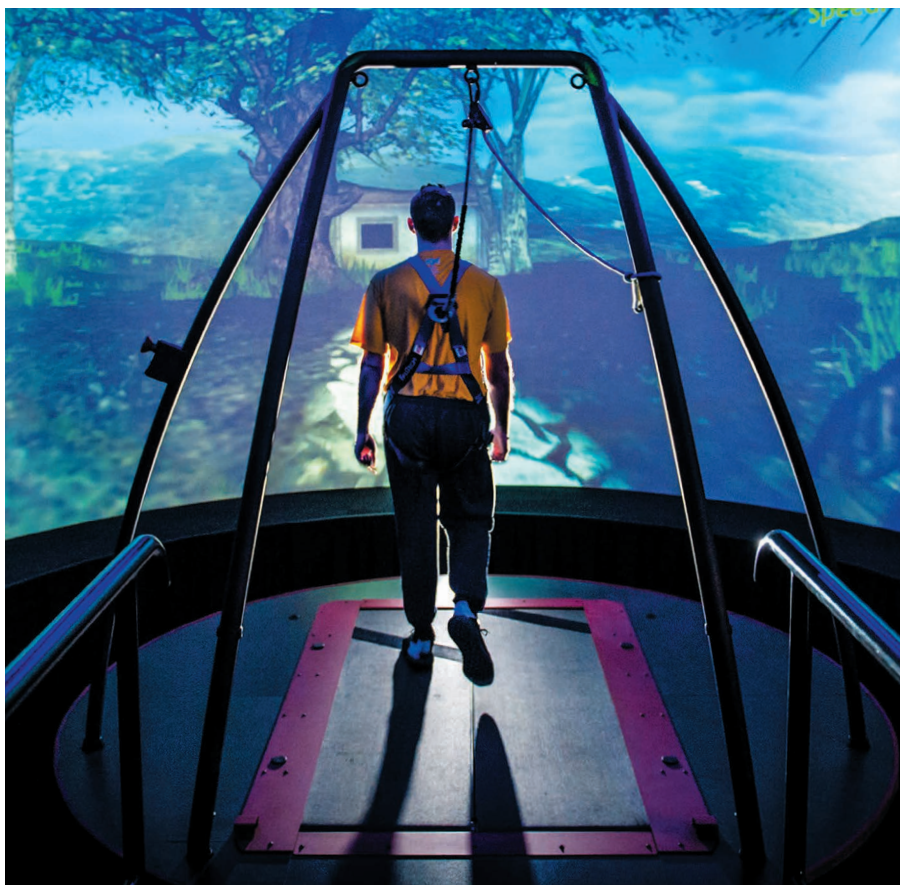
Peer review would change too. Instead of a few reviewers looking at the entire manuscript, several would do so, each focusing closely on a particular substudy. In this way, submissions that use multiple, diverse techniques will get appropriate scrutiny, helping to avoid the publication of papers that are like "grand mansions of straw".

Finally, funders, research institutions and journals would need to explicitly support publication of weightier articles. Or perhaps we need to develop formal ways — beyond simple citations — to explicitly link and recognize substudies that triangulate a single question.

A proposal published early last year advocated for a new category of paper that combines hypothesis-generating work with robust, pre-registered confirmatory studies conducted by qualified independent labs⁹. Papers involving triangulation in a way we propose will clearly often involve considerable work coordinating groups of researchers from different disciplines. Reviewers and tenure committees should find ways to value them appropriately. ■

Marcus R. Munafò is programme lead and **George Davey Smith** is director at the Medical Research Council's Integrative Epidemiology Unit at the University of Bristol, UK.
e-mail: marcus.munaf@bristol.ac.uk

1. Baell, J. & Walters, M. A. *Nature* **513**, 481–483 (2014).
2. Munafò, M. R. et al. *Nature Hum. Behav.* **1**, 0021 (2017).
3. Lawlor, D. A., Tilling, K. & Davey Smith, G. *Int. J. Epidemiol.* **45**, 1866–1886 (2016).
4. Kaelin, W. G. Jr *Nature* **545**, 387 (2017).
5. Chokshi, D. A., El-Sayed, A. M. & Stine, N. W. *J. Am. Med. Assoc.* **314**, 1339–1340 (2015).
6. Krieger, N. & Davey Smith, G. *Int. J. Epidemiol.* **45**, 1787–1808 (2016).
7. Lipton, P. *Inference to the Best Explanation* 2nd edn (Routledge, 2004).
8. Rennie, D., Yank, V. & Emanuel, L. *J. Am. Med. Assoc.* **278**, 579–585 (1997).
9. Mogil, J. S. & MacLeod, M. R. *Nature* **542**, 409–411 (2017).



Virtual reality in use as part of physical therapy after a traumatic brain injury.

TECHNOLOGY

Virtual reality comes of age

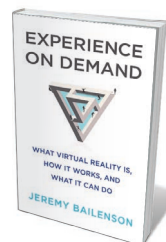
Ramin Skibba weighs up a hymn to the technology's applications, from school to sports.

You strap on the head-mounted display, slip on the gloves, tune your ears to the surround sound — and suddenly you are facing a plank jutting out over an abyss. The depths here are virtual, but not everyone can force themselves to jump.

This is just one program developed by psychologist Jeremy Bailenson to demonstrate the capabilities of virtual reality (VR). As a leading researcher in the field, Bailenson crafts new worlds that feel real, to explore their beneficial uses. In *Experience On Demand*, he tours the myriad applications that he and others are developing. After a great deal of hype by science-fiction film writers and video-game designers in the

1990s, the technology now finally seems poised for widespread use. Eventually, as Bailenson details, it could transform work, schools, hospitals and more.

Fast, high-resolution VR systems such as Bailenson's excel as training tools because they so effectively recreate interaction with a particular environment: a user's motor and perceptual systems interact with the surroundings more or less as they would with the real thing. Psychologists refer to this as "presence"; it is, as Bailenson notes, "the fundamental characteristic of VR". The system tracks your every move, providing a realistically shifting sensory perspective. Small things loom as you move towards them; the view turns as you rotate your head. Props



Experience on Demand: What Virtual Reality Is, How It Works, and What It Can Do

JEREMY BAILENSON
W. W. Norton: 2018.

or a shaking floor can make an experience feel very real indeed.

The applications, as Bailenson details, are legion. VR is an efficient way to train workers in dangerous or challenging jobs. For example, quarterbacks in American football need daily strategy practice alongside their cardio and weights, to prepare for every possible defence. Bailenson

developed a training program for the team at Stanford University in California; he has now expanded it into STRIVR, a company offering immersive training. The firm provides tools that claim to improve performance and boost productivity in a range of companies and sports teams.

VR also lends itself to social, ethical and environmental education. Bailenson discusses examples of how it could be used to tackle ageism and reduce waste. The idea is that it can give users any kind of body. In one of Bailenson's studies, for instance, participants who were given an 'elder' avatar and saw themselves in a virtual mirror showed a 20% improvement against ageist stereotypes in one measure of bias. This was a word-association task posing questions such as "When you think of somebody old, what are the first five words that come to mind?"; people who had experienced the elder avatar used more positive words. However, the tactic backfired with respect to race. White people who tried on a black avatar subsequently scored worse in a test of implicit bias. In this case, rather than boosting empathy, the virtual experience primed racist stereotypes.

Programs can also be used for physical and psychological therapy. People with burn injuries experienced up to 44% less pain when using VR because the immersive environment was such an efficient distraction, according to a study by psychologist Hunter Hoffman at the University of Washington in Seattle and his colleagues (Y. S. Schmitt *et al. Burns* 37, 61–68; 2011). VR has also been used to help people with post-traumatic stress disorder to gradually come to terms with their traumatic experience.

There are inevitable risks and drawbacks. Mayank Mehta, a neurophysicist at the University of California, Los Angeles, investigated the effects of the technology on the brains of laboratory rats. His team found that rats respond to the sight of a virtual dispenser of sugar water as if it is the real thing, running faster towards it and even salivating and licking as they (virtually) approach

it — a sign of addiction (Z. M. Aghajan *Nature Neurosci.* **18**, 121–128; 2015). In a 2014 study by Frank Steinicke and Gerd Bruder at the University of Hamburg in Germany, a participant started blurring the distinctions between real and virtual objects after immersion in a virtual environment many times in a single day (F. Steinicke and G. Bruder *Proc. 2nd ACM Symp. on Spatial User Interaction* 66–69; 2014).

Bailenson mentions escapist, excessive use of VR as a major risk. Because of “simulator sickness” and eye strain, which can develop after just 20 minutes, this has not yet been studied in humans. It is as yet a speculative concern, explored more in film and fiction. In addition, there are concerns that violent programs, such as VR versions of first-person-shooter video games, might encourage antisocial or aggressive behaviour in the real world. But Bailenson gives such concerns short shrift. Nor does he call for transparency or oversight of VR companies, or for regulations to ensure consumers’ safety. He seems confident that developers and users will know how to use the technology responsibly.

Indeed, Bailenson is, by his own admission, “bullish” about VR; he recognizes that he might have “drunk the Silicon Valley Kool-Aid”. That relentless positivity means that the book can lack nuance, as if VR can

“Virtual reality lends itself to social, ethical and environmental education.”

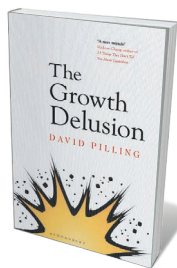
solve the world’s problems. Bailenson, for instance, wants to combat climate change by using the technology to encourage people to change their behaviour, for example by taking shorter showers and making fewer long-distance flights. He also wants to see it used in schools for virtual field trips — although the cost of the equipment would make access unequal.

Social-media trolls pose another problem. Platforms such as Facebook — which acquired the VR company Oculus in 2014 — could one day incorporate virtual interactions, raising the chilling spectre of increasingly realistic goading and abuse.

Bailenson often writes like a scientist. His prose can be verbose, peppered with jargon such as “boundary conditions”. He verges on the grandiose, calling VR a “movement” or a “revolution”. Nevertheless, his enthusiasm is contagious, and he explains complex issues to an audience broader than fellow scientists, providing a real vision of our possibly VR-infused future. ■

Ramin Skibba is an astrophysicist turned science writer based in San Diego, California.
e-mail: raminskibba@gmail.com

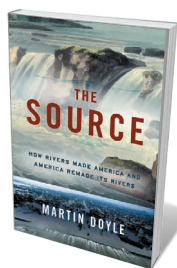
Books in brief



The Growth Delusion

David Pilling BLOOMSBURY (2018)

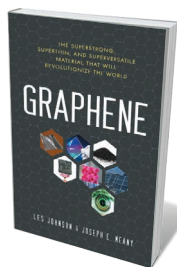
“Only in economics is endless expansion seen as a virtue. In biology it is called cancer.” Rarely does a study of gross domestic product (GDP) and growth sizzle with such wit and acuity, but *Financial Times* editor David Pilling manages the feat. He skewers the linked concepts as a statistical neverland that factors in crime and ignores housework. He pulls out absurdities such as the stratospheric US health-care costs that prop up the nation’s economic well-being, yet destroy uninsured families. And he presents a cogent argument for the multi-index ‘dashboard’ superseding mere GDP. Masterful.



The Source

Martin Doyle W. W. NORTON (2018)

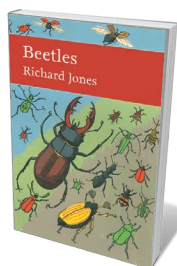
Rivers have shaped the United States geologically, economically and demographically — there are, after all, 250,000 in the country. This history by water-policy expert Martin Doyle nimbly explores that process in tandem with the heroic era of US construction that saw the rise of projects such as the Grand Coulee Dam. In his telling, rivers become a lens on federalism, energy and conservation — a rolling narrative taking us from George Washington’s quest to find a passage from the Atlantic Ocean to the Ohio River, through decades of levee-building, flood control, water wars and much more.



Graphene

Les Johnson and Joseph E. Meany PROMETHEUS (2018)

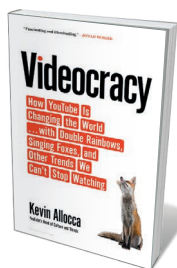
How can a material one atom thick conduct electricity or filter filthy water? Physicist Les Johnson and chemist Joseph Meany tell all about graphene, that wispy “tessellation of carbon atoms” finally coming into its own. Their primer is fittingly slim, but covers an impressive swathe of the science and its applications. Along with a lucid history of earlier carbon “miracle materials”, they follow the path from lab to production. The potential is vast, from making the material using waste carbon dioxide harvested from astronauts’ breath, to creating graphene-based transistors that detect harmful genes.



Beetles

Richard Jones WILLIAM COLLINS (2018)

It’s no surprise that Alfred Russel Wallace and Charles Darwin were both avid fans of the beetle. The nearly half a million described species of Coleoptera are like animated jewels, from their gaudy wing-casings to their shiny, secateur-like mandibles. Entomologist Richard Jones’s illustrated tome (part of the Collins New Naturalist Library) ranges over their anatomy, natural history and behaviour. Things get really wild with the defensive ‘chemical cannon’ of the bombardier beetle, and the biscuit beetle’s reduction of noodles to “ticker tape and dust”. Watch out — there are wonders underfoot.



Videocracy

Kevin Allocca BLOOMSBURY (2018)

YouTube can seem like a parallel universe — a trove of cultural data so huge it would take years to watch the content posted in a day. This ‘biography’ of the web-video behemoth by its trends director, Kevin Allocca, tours the technology and the clips that have trended or gone viral, from astronaut Chris Hadfield singing David Bowie’s ‘Space Oddity’ on board the International Space Station, to Egyptian protests during the Arab Spring. Allocca examines, too, the darker side of mass cultural participation, such as the raising of troll armies. **Barbara Kiser**

Correspondence

China's ban could curb plastic waste

China's ban on imports of recycled plastic from developed countries takes effect this month. It could be a game changer if it weans us off plastic and forces us to seek sustainable alternatives.

With no suitable strategies in place for dealing with this extra unexpected plastic, countries must quickly devise and implement alternative waste-management solutions (see also C. M. Rochman *et al. Nature* **494**, 169–171; 2013). Many jurisdictions have legislation that prohibits the dumping of plastic waste into landfill. And stockpiling plastic refuse is ill-advised, given the fire risk at storage sites (see, for example, go.nature.com/2dh3mbg).

Moves to change consumer behaviour and implement strategies to cut plastic usage are gaining momentum. International policies and financial disincentives meant to curb the proliferation of single-use plastics (plastic bags and microbeads) are already showing positive results (D. Xanthos and T. R. Walker *Mar. Pollut. Bull.* **118**, 17–26; 2017). These should be extended to include a ban on other items, such as plastic drinking straws, and by widely introducing deposit-and-return schemes for plastic bottles.

Tony R. Walker *Dalhousie University, Halifax, Canada.*
trwalker@dal.ca

Top genes: the most common searches

Online genetic information is widely explored by the public as well as by researchers (see *Nature* **551**, 427–431; 2017). To get a sense of which genes attract the most public attention, I used Google Trends to gather statistics from 2004 to the present (see go.nature.com/2dsjvdm).

I found that cancer-related genes are among the most commonly searched. The

top-scoring gene you identify for researchers searching PubMed, *TP53*, also gathered the highest number of queries on Google — as might be expected from the role of mutant p53 proteins in tumour development. Search queries for *BRCA1* peaked when actor Angelina Jolie announced her preventive double mastectomy in 2013 and the removal of her ovaries and fallopian tubes in 2015. And the promising development of cancer immunotherapy has coincided with a surge in queries in the past few years for *PD-L1*, a targeted immune-checkpoint gene.

The growing popularity of genetic testing for mutations that substantially increase the risk of disease has also brought fame to certain genes. Examples include mutations in *APOE* in Alzheimer's disease, in *SERPINA1* in α -1 antitrypsin deficiency, and in *CFTR* in cystic fibrosis. The scientific community can further empower the public through timely and accurate communication of genetic findings.

Kuan-lin Huang *Washington University in St. Louis, Missouri, USA.*

kuan-lin.huang@wustl.edu

Top genes: names confound hit parade

A rough proportionality might be expected between the number of citations a gene collects in PubMed (see *Nature* **551**, 427–431; 2017) and the hits it receives in Internet searches — where the former reflects its scientific value and the latter is also influenced by its impact on the wider public. Sometimes, however, the names of the genes themselves may introduce anomalies that distort this relationship.

Some gene names are much more popular outside science than they are on PubMed. 'Superman' is an example, referring as it does to a cult figure as well as to the *SUPERMAN* gene in the thale

cross *Arabidopsis thaliana*.

This distortion is particularly pronounced for longer gene names that are full words or phrases, such as *drop dead* and *Brokenheart* in the fruit fly *Drosophila melanogaster* (M. R. Seringhaus *et al. Genome Biol.* **9**, 401; 2008).

Moreover, this distortion may be evident for genes that are now rarely a focus in the literature but still attract search-engine hits on a scale comparable to scientifically popular genes such as *TP53*, which encodes the tumour-suppressor protein p53. The gene for alcohol dehydrogenase (*ADH*), the enzyme responsible for metabolizing alcohol, is such an example.

Mark B. Gerstein, Fabio

C. P. Navarro *Yale University, New Haven, Connecticut, USA.*
pi@gersteinlab.org

Virtual carbon price is worth testing

Academic institutions, non-profit organizations and the public sector are experimenting with and sharing their findings on internal carbon pricing (K. Gillingham *et al. Nature* **551**, 27–29; 2017). They are also well positioned to experiment with proxy (or shadow) carbon prices as decision-making tools.

Proxy carbon pricing incorporates a virtual carbon tax into a financial decision without collecting revenue. It can be applied in selective ways — for example, to only the largest capital investments or to particular kinds of purchasing (see, for instance, go.nature.com/2dparje).

Several questions need to be collectively addressed. It is unclear how planning and purchasing processes can best be altered to incorporate proxy carbon prices. Effective use of a proxy price usually requires a life-cycle cost assessment, which is not standard practice in many institutions. Researchers also do not fully understand the institutional and technical

conditions that cause a proxy carbon price (as opposed to energy savings) to alter a business decision.

A broad range of organizations need to share their findings from the use of these tools, including how they interact with other decision criteria.

Alexander R. Barron, Breanna J. Parker *Smith College, Northampton, Massachusetts, USA.*

abarron@smith.edu

Regulate prescription of Chinese medicines

Problems with prescriptions for traditional Chinese medicines (TCMs) threaten to create a chasm between the Chinese government's medical reforms and their outcome (see *Nature* **551**, 552–553; 2017).

Prescriptions for TCMs are unsupervised in China. Excessive amounts have long been prescribed for clinicians' financial gain. They often involve high-risk injections of unknown efficacy. And some 70% of TCMs are prescribed by untrained practitioners, with almost half of all prescribed medicines proving ineffective (see go.nature.com/2doqbcz; in Chinese).

Correct diagnosis is essential for successful treatment with TCMs. If the indications on the label are used as the only guide, TCMs will not improve patients' health and may even aggravate their conditions.

In addition to improving the quality of TCMs, the Chinese government must regulate prescription practices. It should set up a prescription-review system to prevent misuse and ensure that all clinicians are formally trained in TCM.
Zhijie Xu *Naval Medical University of the Chinese People's Liberation Army, Shanghai, China.*
Lizheng Fang *Zhejiang University, Hangzhou, China.*
Dingzhi Pan *Shanxi Grand Hospital, Taiyuan, China.*
aiolos1025@163.com

Calestous Juma

(1953–2017)

International-affairs scholar who championed science for African development.

“Africa,” Calestous Juma wrote to me in 2015, “is diverging between those who want to talk and those who want to do something practical.” Juma was one of the latter. An international-development scholar, he championed the harnessing of science, technology and innovation for development. He founded Africa’s first science-policy think tank, led major United Nations science initiatives and wrote influential books. Juma, a Kenyan professor at the Harvard Kennedy School’s Belfer Center for Science and International Affairs, died in Cambridge, Massachusetts, on 15 December, at the age of 64.

Juma’s trademark mix of candour and humour inspired many African presidents, including Paul Kagame of Rwanda, to invest in national and continental research schemes. For African academics, Juma was an ally connected to the world’s most powerful presidents and prime ministers. Yet he was loved for his approachability — especially by journalists such as me, with whom he shared a special bond.

Born in 1953, Juma grew up in Busia County, western Kenya, on the shores of Lake Victoria. His childhood was plagued by bouts of malaria. To help pay his school fees, Juma fixed broken radios and record players. Unable to afford university, he trained as a science teacher, but got a job reporting on science and the environment after an editor at Kenya’s *Daily Nation* spotted his exceptional talent for writing in letters Juma submitted to the newspaper. In 1979, he went to work for the non-governmental organization Environment Liaison Centre, based in Nairobi, as a researcher and editor. He went on to receive a scholarship to study science policy at the University of Sussex, UK, where he completed his PhD in 1987.

Juma returned to Kenya to create the African Centre for Technology Studies (ACTS) in Nairobi. ACTS, which opened in 1988, helped to draft Kenya’s first industrial-property legislation, leading to the creation of the country’s patent office. At ACTS, Juma directed a Canada-funded project called Economic Reform and Environment in Africa, which explored the links between economic development and conservation management. Drawing on a three-year project in Africa, he published *The Gene Hunters* (Princeton Univ. Press) in 1989, which set



out the threats of modern biotechnology and its potential for solving food insecurity, especially in developing nations.

In 1995, Juma moved to Canada to serve as the first executive director of the United Nations Convention on Biological Diversity, which he helped to negotiate with policy bodies such as the Food and Agriculture Organization of the UN. He enjoyed engaging scholars, diplomats and researchers in discussions about conservation and sustainable biodiversity. The resulting international agreement on handling the products of modern biotechnology, the Cartagena Protocol on Biosafety, was adopted in 2000, after Juma left the organization. Juma felt that it placed too many restrictions on the use of genetically modified crops in Africa.

In 1998, Juma moved to Harvard to think and write. He spent the early 2000s coordinating a UN task force on how science and technology could assist with the attainment of the Millennium Development Goals, notably eradicating hunger and ensuring environmental sustainability. He influenced Africa’s 2005 science plan, which created continental schemes to boost research under the auspices of the African Union and the New Partnership for Africa’s Development in Midrand, South Africa. One of its fruits is the Southern Africa Network for Biosciences, an initiative based in Pretoria that provides African researchers with access to world-class labs for work on agriculture and health.

In 2007, Juma was the keynote speaker at the first African Union summit that had a focus on science and technology. He urged the heads of state, gathered in Addis Ababa, Ethiopia, to harness knowledge to help their

countries leapfrog industrialized nations.

Juma’s optimism and appetite for action was at odds with the lumbering bureaucracy of African policymaking. He was often frustrated with the slow pace of implementation, and it irked him that science and technology policies were drawn up separately from relevant economic, industrial and social-development policies. He rejected the view that science could drive development through targeted calls from funding agencies for proposals from academics in ivory towers. Rather, he believed in training young Africans to be entrepreneurs and engineers, by investing in infrastructure such as roads and broadband networks and unlocking

African curiosity and ingenuity. “Really, I’m just a cheerleader for African leaders and youth,” he told the *Huffington Post* in 2014.

Juma was no stranger to controversy. His support for biotechnology in developing countries saw him lock horns with people who were lobbying against genetically modified organisms. His book *Innovation and Its Enemies* (Oxford Univ. Press, 2016) charted the battle between “innovation and incumbency” throughout human history. He showed how the fears that led people to initially reject novelties such as coffee, margarine and printing rarely came to pass.

Juma leaves a lasting legacy, not least through the people he met and inspired with his inquisitiveness and mischievous approach. His graduate courses at Harvard on the role of innovation in economic growth and the global economic impacts of biotechnology were popular — in part because of his entertaining lecturing style. True to his vision of getting academic thought out into the real world, he also taught an executive course for senior policymakers and practitioners on how to integrate science and technology into national development policies.

Juma was modest about his achievements, and sanguine about failure, both his own and others’. Development, he maintained, was by its nature experimental, and Africans must be allowed to experiment — to make mistakes, and to learn from them. ■

Linda Nordling is a science writer based in Cape Town, South Africa. She knew Juma from 2005, in her role as founding editor of *Research Africa*, an online research-policy news portal. She tweets as @lindanordling.

MARTHA STEWART

APPLIED PHYSICS

Trapped particle makes 3D images

A technique in which a small particle is trapped and moved by laser light has been used to produce visual representations of objects in three dimensions, offering key advantages over currently used approaches. [SEE LETTER P.486](#)

BARRY G. BLUNDELL

Devices known as volumetric displays allow 3D images to be generated in a transparent enclosure. Because these images occupy three dimensions, they exhibit the spatial characteristics that we associate with real-world scenes. The images can be viewed without the need for glasses by many simultaneous observers, and changes in vantage point allow content to be seen from different orientations. On page 486, Smalley *et al.*¹ describe an innovative approach to volumetric-display implementation that allows 3D images to be formed in the air, removing the need for a transparent enclosure.

For more than 100 years, volumetric displays have been the subject of extensive research². Although it is relatively easy to make a small (tabletop) display that works fairly well, it is extremely difficult to develop a larger display that works very well. There are two overarching (but often conflicting) problems. The first relates to the techniques that are currently used to produce dynamic images of relatively high visual quality. The second concerns the optical characteristics of the imaging volume, which must allow light emanating from the image to propagate, and emerge from the volume, without distortion — think of the distortion that occurs when light emerges from a tropical-fish tank.

With respect to the first problem, in most volumetric displays, the imaging volume is formed by the cyclic motion of a transparent surface (Fig. 1a). To produce a 3D image, a sequence of image slices is depicted on the surface as it moves through the volume. Given the need to refresh images at least 30 times per second to avoid perceptible flicker³, the surface must move rapidly.

The motion of the surface can be either translational (along a straight line) or rotational. When translational motion is used, the dimensions of the imaging volume are limited by mechanical issues arising from the surface's mass and acceleration. In the case of rotational motion, the surface's linear speed increases with distance from the axis of rotation. This impinges on image quality and so can ultimately restrict the diameter of the imaging volume. There is also a 'dead' region

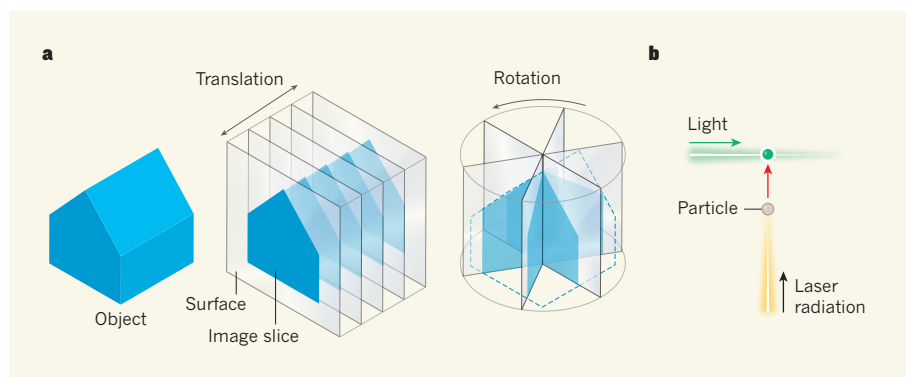


Figure 1 | Volumetric-display techniques. **a**, Devices known as volumetric displays can produce visual representations of objects in three dimensions. They typically use the rapid, cyclic motion of a transparent surface. To generate a 3D image, a sequence of image slices is depicted on the surface as it moves. This motion can be either translational (along a straight line) or rotational. **b**, Smalley *et al.*¹ report an alternative approach in which non-visible laser radiation is used to move a small particle (red arrow). To create an image point, the particle is illuminated with light as it passes through the required position.

in the vicinity of the rotational axis, in which image points cannot be formed⁴.

A further limitation of these displays is that the surface's movement precludes the insertion of haptic probes — tools that recreate the sense of touch by applying forces, motion or vibrations to the user. Such probes can simulate the solidity associated with physical versions of images, so that, for example, virtual clay could be moulded and would feel like real clay.

Smalley *et al.* sought to overcome all of these difficulties using the photophoretic effect⁵, whereby laser light is used to trap and move small particles (with diameters of 5–100 micrometres). To create a point of light at a given location in 3D space, the authors used non-visible laser radiation to move a particle, and as the particle passed through the required position, it was illuminated with red, green or blue light (Fig. 1b). The authors suggest that complex, high-fidelity, dynamic images could be formed by introducing parallelism — the simultaneous movement of many particles.

There are at least three key advantages of Smalley and colleagues' approach. First, it does not require the cyclic motion of a surface — movement is restricted to that of low-mass particles. Second, the presence of these particles will have minimal impact on the propagation of light through the imaging volume. And third, because the image is formed in the

air, image components can coexist with haptic probes and other interaction tools.

The authors provide several photographs of image content produced using their technique (see Figure 2 of the paper¹). However, these photographs required long exposure times — of the order of tens of seconds. For implementing a viable display, there is therefore a pressing need to explore ways of increasing the speed of particle motion and of introducing parallelism such that many image points can be created simultaneously.

The introduction of a high degree of parallelism poses a further challenge, relating to the fact that each point in the imaging volume must be individually accessible. This is reminiscent of an equivalent problem that was encountered in the late 1960s, in connection with a type of 3D display called a photochromic-based volumetric display^{6,7}. Another concern is that the insertion of haptic probes into the image volume will probably give rise to shadow regions that will interfere with the propagation of light used for particle motion and illumination. However, the judicious design of such probes would ameliorate this potential problem.

In terms of photorealism, it is unlikely that these devices will ever directly compete with high-end stereoscopic 3D displays. However, despite more than a century of research into volumetric displays, there has been relatively

little work on exploring ways of capitalizing on key image characteristics. In particular, volumetric displays provide considerable freedom in viewing position, and support both vertical and horizontal motion parallax, which means that observers can move and change their view of an image in a wholly natural way.

Consequently, these devices offer exciting, and largely unexplored, opportunities to advance spatial imaging (in areas such as neurosurgery) and dynamic imaging (in fields including fluid dynamics, robotics and sports training). With regard to the latter, there is a need to better support the visualization of

complex forms of 3D motion⁸. Moreover, creating volumetric images in the air enables direct interaction, thereby allowing, for example, 3D design tasks to be carried out in a natural way in 3D space.

Smalley and colleagues' approach could provide the foundation for the next generation of volumetric displays. Such devices will not only enhance our understanding of complex spatial and geometric dynamics, but also support innovative user interaction. ■

Barry G. Blundell is with the University of Derby Online Learning (UDOL), University of

Derby, Derby DE22 1GB, UK.
e-mail: barry.blundell@physics.org

1. Smalley, D. E. *et al.* *Nature* **553**, 486–490 (2018).
2. Luzy, E. & Dupuis, C. Procédé pour obtenir des projections en relief. French Patent 461,600 (1914).
3. Blundell, B. G. *Enhanced Visualization: Making Space for 3-D Images* 68 (Wiley, 2007).
4. Blundell, B. G. & Schwarz, A. *Volumetric Three-Dimensional Display Systems* 72–91 (Wiley, 2000).
5. Davis, E. J. *Aerosol Sci. Technol.* **26**, 212–254 (1997).
6. Adamson, A. W. Method and apparatus for generating three-dimensional patterns. US Patent 3,609,706 (1971).
7. Adelman, A. H. & Lewis, J. D. Method and apparatus for generating three-dimensional patterns. US Patent 3,609,707 (1971).
8. Blundell, B. G. *3D Res.* **8**, 11 (2017).

STRUCTURAL BIOLOGY

Ageing-related receptors resolved

Ageing is a regulated process in which hormones have pivotal roles. Crystal structures of two hormone co-receptors should be informative for drug discovery focused on age-related disorders. [SEE ARTICLE P.461 & LETTER P.501](#)

MAKOTO KURO-O

In Greek mythology, three goddesses known as the Fates govern the lifespan of each person. Klotho, Lachesis and Atropos are the spinner, the allotter and the cutter of the thread of life, respectively. So when a genetic mutation was identified in mice that undergo premature ageing¹, the gene involved was fittingly named *klotho*. The protein it encodes, α -klotho, and a sister protein called β -klotho, are high-affinity co-receptors for certain members of the fibroblast growth factor (FGF) family of signalling proteins², but their means of action has not been well characterized. Two papers^{3,4} in this issue describe crystal structures of FGF-klotho complexes, not only providing a basis for understanding how klothos act, but also opening up avenues for structure-based drug design.

α -Klotho is a membrane-spanning protein expressed predominantly in the kidney, as well as in the brain. Mice lacking α -klotho exhibit a range of signs associated with ageing, including hearing loss, impaired cognition and organ atrophy⁵. They also have elevated blood phosphate levels. However, the protein's function on the molecular level was unclear, until mice lacking FGF23 were characterized⁶.

FGF23 is one of the three endocrine FGFs, which act as hormones, secreted by one organ to regulate the function of another. Specifically, FGF23 is secreted from bones after phosphate intake and acts in the kidney to inhibit phosphate reabsorption in urine, thereby maintaining the body's phosphate balance. Mice lacking FGF23 have elevated phosphate levels

owing to impaired phosphate excretion, and exhibit features associated with ageing⁶. This striking similarity to mice lacking α -klotho led researchers to discover² that α -klotho forms a complex with the membrane-spanning protein FGF receptor 1c (FGFR1c), acting as a co-receptor to recruit FGF23 and so triggering FGF signalling.

In the first of the current studies, Chen *et al.*³ (page 461) solved the crystal structure of FGF23 in complex with the ligand-binding domain of FGFR1c and the extracellular domain of α -klotho. The structure revealed that α -klotho (aptly, given its namesake) sends out a long receptor-binding arm (RBA) that

acts as a thread to capture the ligand-binding domain of FGFR1c. Indeed, when the authors generated α -klotho lacking the RBA, the mutant protein failed to capture FGFR1c or to help FGF23 to activate FGF signalling.

Chen and colleagues showed that FGF23 fits into the groove created between α -klotho and FGFR1c. The globular amino-terminal region and the rod-like carboxy-terminal region of FGF23 face FGFR1c and α -klotho, respectively (Fig. 1). By promoting formation of this complex, α -klotho enables strong interactions between FGF23 and FGFR1c, which otherwise interact only weakly.

Like α -klotho, β -klotho functions as a co-receptor for endocrine FGFs, forming a complex with FGFR1c to bind FGF21, and with FGFR4 to bind FGF19 (refs 7,8). FGF19 is secreted from the intestine after feeding, and acts in the liver to suppress bile-acid synthesis. FGF21 is secreted from the liver following fasting, and acts in fat cells and the brain to induce metabolic adaptation to fasting and responses to stress⁵. Although FGFRs are expressed in a wide range of tissues, the tissue-specific expression of β -klotho in the liver, fat and brain restricts the target organs of these endocrine FGFs.

In the second study, Lee *et al.*⁴ (page 501)

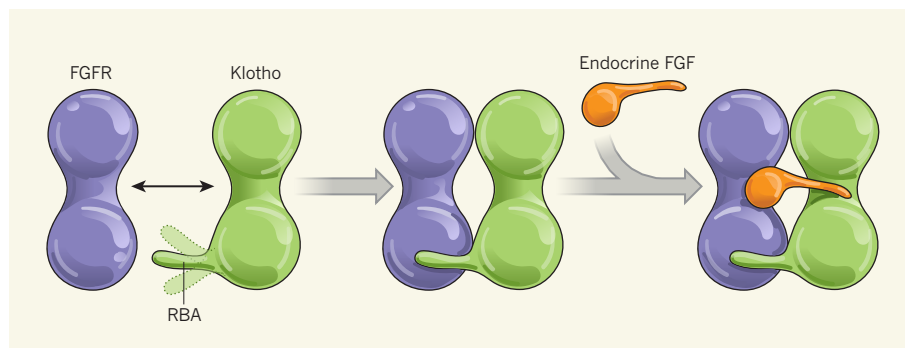


Figure 1 | Structures reveal the mode of action for klotho proteins. Two groups^{3,4} have produced crystal structures of the extracellular domains of klotho proteins, either alone, in complex with 'endocrine' fibroblast growth factors (FGFs), or in complex with both endocrine FGFs and the ligand-binding domains of FGF receptors (FGFRs). As this simplified schematic shows, the klotho proteins seem to have an intrinsically disordered receptor-binding arm (RBA) with which they capture FGFRs (interaction indicated by double-headed arrow). The RBA enables formation of a stable complex, with FGFs fitting into the groove between the other two proteins.

resolved the crystal structure of β -klotho's extracellular domain when bound to and when free from FGF21, in the absence of FGFRs. Like FGF23, the C-terminal region of FGF21 fits into the groove in β -klotho. However, the authors could not solve the structure of some regions in β -klotho, including that corresponding to the RBA in α -klotho. This suggests that the RBAs of klotho proteins are intrinsically disordered and unable to fold stably unless bound to FGFRs. The fact that intrinsically disordered proteins can interact with multiple proteins⁹ implies that the RBAs of klotho proteins could capture other partners besides FGFRs. This might explain why the extracellular domain of α -klotho, which can be released into the extracellular space, has been reported to have FGF-independent activity, regulating several ion channels and transporters, along with other growth factors and their receptors⁵.

Another proposed FGF-independent activity for the klothos is as carbohydrate-binding proteins called lectins. Klothos belong to a family of enzymes that cut sugar chains¹, but not all of the amino-acid residues essential for this enzymatic activity are found in the klothos. Thus, klothos might bind to, but not cut, specific carbohydrates. Lee and colleagues' structure of β -klotho leaves open the possibility that this protein interacts with particular sugar chains. By contrast, Chen and co-workers' structure of FGF23- α -klotho-FGFR does not fit with the idea of α -klotho acting as either an enzyme or a lectin. However, it might be that in the absence of FGFRs, the structure of α -klotho would provide a different point of view. Alternatively, it is possible that the two klothos have different FGF-independent activities.

FGF-klotho signalling has key roles in ageing and age-related disorders. The new structures could be used to develop drugs to treat disorders of ageing, using structure-based drug design to identify targets in FGF-klotho-FGFR complexes. For instance, consider chronic kidney disease (CKD)^{5,10} — a common state of impaired renal function that often occurs as a complication of high blood pressure or diabetes. People with CKD exhibit many of the same symptoms as mice lacking α -klotho, including disturbed phosphate metabolism and increased risk of death^{5,10}. Placing mice lacking FGF23 or α -klotho on a low-phosphate diet reduces the phosphate retention and premature ageing normally seen in these animals, indicating that phosphate increases accelerate ageing¹⁰. Thus, drugs that target FGF-klotho-FGFR complexes to improve phosphate metabolism might be useful to treat CKD.

A second example lies in the targeting of FGF21 complexes. FGF21 overexpression extends lifespan in mice¹¹, and this protein has been dubbed an 'anti-ageing' hormone. Lee *et al.* demonstrated that they could increase the potency of FGF21 by introducing genetic

mutations designed to increase the protein's affinity for β -klotho. Further analyses such as this could provide a way to explore anti-ageing medicines more generally. ■

Makoto Kuro-o is in the Division of Anti-Ageing Medicine, Center for Molecular Medicine, Jichi Medical University, Shimotsuke, Tochigi 329-0498, Japan. e-mail: mkuroo@jichi.ac.jp

1. Kuro-o, M. *et al.* *Nature* **390**, 45–51 (1997).
2. Kurosu, H. *et al.* *J. Biol. Chem.* **281**, 6120–6123 (2006).

3. Chen, G. *et al.* *Nature* **553**, 461–466 (2018).
4. Lee, S. *et al.* *Nature* **553**, 501–505 (2018).
5. Hu, M. C., Shiizake, K., Kuro-o, M. & Moe, O. W. *Annu. Rev. Physiol.* **75**, 503–533 (2013).
6. Shimada, T. *et al.* *J. Clin. Invest.* **113**, 561–568 (2004).
7. Ogawa, Y. *et al.* *Proc. Natl Acad. Sci. USA* **104**, 7432–7437 (2007).
8. Kurosu, H. *et al.* *J. Biol. Chem.* **282**, 26687–26695 (2007).
9. Wright, P. E. & Dyson, H. J. *J. Mol. Biol.* **293**, 321–331 (1999).
10. Kuro-o, M. *Nature Rev. Nephrol.* **9**, 650–660 (2013).
11. Zhang, Y. *et al.* *eLife* **1**, e00065 (2012).

This article was published online on 17 January 2018.

ASTRONOMY

A beacon at the dawn of the Universe

Quasars are the brightest continuously emitting sources of radiation in the Universe. Measurements of the most distant quasar ever detected reveal details about the evolution and structure of the early Universe. SEE LETTER P.473

EILAT GLIKMAN

Since their discovery¹ in 1963, astronomical objects called quasars have been among our most powerful probes of the early Universe. Initially seen as mysterious sources of extreme luminosity, quasars are now known to be supermassive black holes that are voraciously consuming gas from their immediate surroundings, emitting large amounts of radiation in the process. On page 473, Bañados *et al.*² report observations of the most distant quasar found so far. The light detected from this object was emitted when the Universe was

a mere 690 million years old — just 5% of its current age.

Almost 90 years ago, the astronomer Edwin Hubble discovered that the Universe is expanding³. The expansion stretches light waves travelling through space, such that light that was emitted from a distant source as blue might be detected as red. This phenomenon is called redshift, and is associated with both distance and time: the larger the redshift, the farther away the source was when it emitted its light, meaning that the light was emitted at an earlier time.

If we rewind the expansion, we find that the

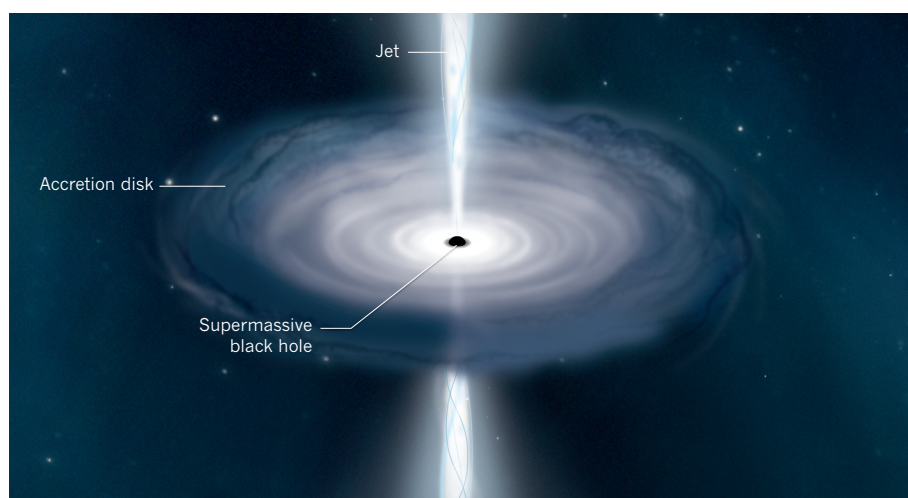


Figure 1 | Emission from a quasar. Quasars are extremely luminous astronomical objects that comprise a supermassive black hole surrounded by an orbiting disk of gas called an accretion disk. As material in the disk is pulled towards the black hole, energy is released in the form of electromagnetic radiation and, in some cases, as beams of charged particles called jets. Bañados *et al.*² report observations of the most distant quasar identified so far, the light of which was emitted when the Universe was only 5% of its current age.

Universe started out in a hot, dense state, filled mostly with ionized hydrogen. As it expanded, it also cooled, and after about 380,000 years, the temperature was low enough for neutral hydrogen to form. For the first few hundred million years, the Universe was devoid of any sources of light — no stars, galaxies or quasars existed. The first stars were then born, but the Universe remained dark because neutral hydrogen is highly effective at absorbing ultraviolet radiation (the main type of emission from these stars).

However, the present-day Universe is filled with sources of light, and the hydrogen that exists in the space between galaxies (the intergalactic medium) is completely ionized and therefore transparent to the ultraviolet emission from early galaxies and quasars. The process of this phase change from a neutral to an ionized Universe, known as reionization, is poorly understood.

The neutral fraction of hydrogen in the Universe can be estimated by analysing the absorption of light by hydrogen in quasars. Studies of quasars observed as they were when the Universe was 0.85 billion to 1.2 billion years old (corresponding to redshifts of 6.5 to about 5, respectively) have shown that the neutral fraction decreased sharply from 0.1% to 0.01% during this time⁴. However, most of the reionization process occurred before this epoch.

Bañados and colleagues' quasar, known as ULAS J1342+0928, has a redshift of 7.54. This means that its strong ultraviolet emission has been shifted into the near-infrared, beyond the sensitivity of typical imaging surveys of the sky. Finding such a high-redshift quasar was not possible until about a decade ago, when sufficiently sensitive near-infrared detectors began scanning large areas of the sky^{5,6}. By studying the absorption spectrum of ULAS J1342+0928 (the fraction of incident radiation absorbed by the intergalactic medium over a range of frequencies), the authors determined that the neutral proportion of hydrogen was at least 10% when the Universe was 690 million years old, which sets a strong constraint on how the intergalactic medium was reionized.

The quasar's black hole is extremely massive — about 800 million times the mass of the Sun. Black holes grow by consuming (accreting) gas from a surrounding structure called an accretion disk (Fig. 1). The gas emits radiation as it falls in. However, such systems have a maximum luminosity, which occurs when the pressure of the emitted light pushes away the infalling gas, halting further growth. This luminosity depends on the mass of the accreting black hole, and therefore defines a maximum growth rate, known as the Eddington limit, for the system.

Bañados *et al.* suggest that the large mass of the black hole in ULAS J1342+0928 can be explained if the object began its life as an initial (seed) black hole of at least 1,000 solar masses.

This result could rule out models in which black-hole seeds were created from the deaths of the first massive stars⁷, and instead favour models in which these seeds formed from the direct collapse of primordial gas⁸. In addition, the black hole would need to have grown continuously (and, therefore, exponentially) at the Eddington limit, starting from when the Universe was roughly 65 million years old. Although this scenario is physically possible, it requires extreme, sustained accretion for about 600 million years, which is substantially longer than the typical lifetime of a quasar⁹.

So far, only two quasars with redshifts greater than 7 have been discovered. The previous record holder was reported¹⁰ in 2011, and early models of quasar evolution predicted that more should have been found by now¹¹. The methods for finding quasars, even at these high redshifts, are sound and have been proved effective. Therefore, the dearth of high-redshift quasars might indicate that these objects were uncommon in the early Universe, and could imply a sharp decline in quasar activity towards early times¹². If so, this suggests that we might be observing extremely rare systems as they were beginning to emerge in the Universe.

The authors' work offers a glimpse into the conditions of the intergalactic medium at the

earliest epoch of structure formation in the Universe, and could place key constraints on cosmological models of this era. However, a single quasar is insufficient for providing a complete picture of the Universe in the reionization era or of the evolution and growth of supermassive black holes from initial seeds. The task ahead is, then, to mine the upcoming near-infrared sky surveys for additional quasars that can paint a more complete picture of the rapidly evolving early Universe. ■

Eilat Glikman is in the Department of Physics, Middlebury College, Middlebury, Vermont 05753, USA.
e-mail: eglikman@middlebury.edu

- Schmidt, M. *Nature* **197**, 1040 (1963).
- Bañados, E. *et al.* *Nature* **553**, 473–476 (2018).
- Hubble, E. *Proc. Natl Acad. Sci. USA* **15**, 168–173 (1929).
- Fan, X., Carilli, C. L. & Keating, B. *Annu. Rev. Astron. Astrophys.* **44**, 415–462 (2006).
- Wright, E. L. *et al.* *Astron. J.* **140**, 1868–1881 (2010).
- Lawrence, A. *et al.* *Mon. Not. R. Astron. Soc.* **379**, 1599–1617 (2007).
- Natarajan, P. & Volonteri, M. *Mon. Not. R. Astron. Soc.* **422**, 2051–2057 (2012).
- Bromm, V. & Loeb, A. *Astrophys. J.* **596**, 34–46 (2003).
- Hopkins, P. F. & Hernquist, L. *Astrophys. J.* **698**, 1550–1569 (2009).
- Mortlock, D. J. *et al.* *Nature* **474**, 616–619 (2011).
- Fan, X. *et al.* *Astrophys. J.* **121**, 54–65 (2001).
- McGreer, I. D. *et al.* *Astrophys. J.* **768**, 105 (2013).

SUSTAINABILITY

Satellite images show China going green

Large-scale tree-planting projects have taken place in regions of China prone to soil erosion. Satellite imagery reveals the effects of this work, and shows that a predicted vegetation decline didn't occur during a period of drought.

MARC MACIAS-FAURIA

The effects of human activities on Earth's vegetation have tended to be negative, mostly because of deforestation¹. Restoration efforts are often restricted to small, localized scales. Large ecological-engineering projects aimed at producing regional-scale effects are few, and among these, China's mega-projects — most notably, the Grain for Green Project (GGP)² — stand out because of their unparalleled scale (27.8 million hectares of forest re-established as of 2013 across 26 Chinese provinces³). Writing in *Nature Sustainability*, Tong *et al.*⁴ report that the positive effects of these tree-planting projects on vegetation growth can be detected using remote-sensing satellite imagery of a large region of southwestern China (the provinces of Guizhou, Guangxi and Yunnan), in an area

associated with highly erodible landscapes called karst. The authors note that these projects, which require considerable investment, will be justified only if the modification of ecosystem properties can be achieved on a large scale.

The government-run GGP, intended to halt soil erosion and desertification, began in 1999 (ref. 2). The project's goal was to convert land on mountainous terrain prone to erosion (cropland or scrubland) into forested landscapes (Fig. 1). Such forest would be classified as ecological if trees might eventually be logged (subject to permission) as part of a timber quota, and as economical if it contained orchards, or plantations of trees for medical use. Ecological forest accounted for 80% of the planting area, with economical forest making up the remaining 20% (ref. 4). The GGP was developed partly in response to



Figure 1 | Trees planted as part of an ecological-engineering project in China. Tong *et al.*⁴ report an analysis of the effects of a large-scale tree-planting project, called the Grain for Green Project, on mountainous regions of southwestern China that are associated with high levels of soil erosion. Shown here are some trees planted as part of this project in the Wolong Nature Reserve in the southwestern Sichuan province.

the consequences of land-use changes during the time of Chairman Mao Zedong, notably the huge areas logged to provide fuel and construction materials during the Great Leap Forward programme, and large-scale conversions of often marginal, sloping land to agricultural use in the 1960s and 1970s to enhance local self-sufficiency — a change that caused severe erosion problems⁵.

To assess the effects of the tree-planting projects, Tong and colleagues use three independent lines of evidence, and the consistency of the findings convincingly demonstrate the robustness of their results. One approach was the analysis of two complementary properties of vegetation. Satellite-imaging data from 1982 to 2015 allowed the researchers to measure the area of vegetation cover present per square metre of ground (known as the leaf-area index). Other satellite data collected between 1992 and 2012 enabled the authors to assess plant biomass in units of above-ground carbon biomass. Plant biomass can be inferred by converting vegetation optical depth (a property captured by microwave observations that are sensitive to the water content of vegetation) to total carbon using an approach based on the carbon density of above-ground, living, woody vegetation.

Over time, both of these properties revealed a marked transition in the amplitude and/or direction of vegetation trends around the main implementation period of the tree-planting project, between 2000 and 2006. The authors'

calculations indicate that the southwestern region of China that they studied acted as a carbon sink after the GGP implementation, providing a considerable amount of the entire country's net carbon sequestration. The authors also observed negative vegetation trends in the provinces' growing urban areas, such as in the cities of Kunming and Nanning. This provides an indirect validation of the team's satellite-data approach.

The second line of investigation taken by Tong *et al.* involved the use of dynamic ecosystem modelling to explore what might have happened in the absence of the tree-planting project. The model took into account the effect of the increase in atmospheric carbon dioxide on vegetation during the time frame studied. This modelling exercise highlighted the divergence between the simulated trend of vegetation decrease projected if the tree-planting intervention had not occurred — linked to a long-lasting drought during the previous decade — and the vegetation increase that was observed.

The third approach taken by the researchers was an analysis of the number of hectares on which tree-planting actions were implemented in each of the 295 counties within the 3 provinces studied. These GGP-inventory data showed a correspondence between actions at the county level and positive vegetation trends, as well as stark differences between China's provinces and the neighbouring countries of Laos, Vietnam and Myanmar, in which

the vegetation assessed by satellite imagery decreased over the same period.

Tong and colleagues' results are encouraging in regard to the large-scale effects of the GGP on vegetation, but should not be taken as a proof of its overall success. As the authors mention, the satellite trends were not validated by measurements taken on the ground. No erosion assessment was undertaken, so one of the main GGP goals was not evaluated directly. Furthermore, the time span of satellite analysis, and of the programme itself, might still preclude the detection of long-term dynamics related to the long lifespan of trees, or might not take into account the role of large but infrequent erosion or disturbance events such as those linked to torrential rains or pest outbreaks.

Most crucial for the overall assessment of the success of the GGP as an ecological restoration project is the fact that satellite data do not distinguish biological composition, such as the presence of different species, and so cannot be used to assess the project's effects on biodiversity. The GGP focused on the planting of non-native, fast-growing monocultures, which might render the resulting forests more vulnerable to pests^{3,4}. The GGP thus used a narrow view of ecosystem services (the role of vegetation in reducing erosion and desertification rates) that had the additional (and possibly unplanned) benefit of a net carbon-storage outcome. Furthermore, the rationale for GGP actions was based not on previous ecological states or projected overall ecological benefits, but on the potential to reduce the erosion rates on the target land and for the programme to generate income for farmers².

In the absence of a China-wide assessment of the GGP's environmental and ecological impacts², an analysis of data from China based on 258 publications³ identified limited biodiversity benefits of the GGP. This was mainly because the dominant non-native, fast-growing monoculture plantations were linked to a decrease in floral diversity, associated with bee and bird population declines, as observed in Sichuan province. This report³ strongly recommends using native trees when establishing plantations, or at least the establishment of plantations composed of several tree species.

Nevertheless, Tong and colleagues' work clearly shows a large-scale effect of the GGP on vegetation in southwestern China. This important result needs to be complemented by ground-based studies. Understanding of the GGP's functional and biodiversity effects is needed to assess its success, and might also identify other interventions that have the potential to enhance or generate wider positive effects of the GGP as an ecological-restoration mega-project. The task set out by Tong and colleagues for how the effects of such massive initiatives can be tested on an adequate scale is valuable and very welcome. ■

Marc Macias-Fauria is at the School of Geography and the Environment, University of Oxford, Oxford OX1 3QY, UK.

e-mail: marc.maciasfauria@ouce.ox.ac.uk

1. Baccini, A. et al. *Science* **358**, 230–234 (2017).

2. Delang, C. O. & Yuan, Z. *China's Grain for Green*

Program: A Review of the Largest Ecological Restoration and Rural Development Program in the World (Springer, 2015).

3. Hua, F. et al. *Nature Commun.* **7**, 12717 (2016).

4. Tong, X. et al. *Nature Sustain.* **1**, 44–50 (2018).

5. Harkness, J. *China Q.* **156**, 911–934 (1998).

This article was published online on 22 January 2018.

In Retrospect

Eighty years of superfluidity

In 1938, two studies demonstrated that liquid helium-4 flows without friction or viscosity at temperatures close to absolute zero. The finding led to major advances in our understanding of low-temperature physics.

WILLIAM P. HALPERIN

In the early twentieth century, scientists discovered the non-intuitive phenomena of superconductivity and superfluidity, in which electrons and atoms, respectively, flow without resistance over great distances. Superfluidity was beautifully demonstrated 80 years ago in two papers published in *Nature* by Allen and Misener¹ and Kapitza². The authors observed the flow of liquid helium-4 through extremely narrow channels and showed that the substance becomes a superfluid at very low temperatures. The studies presaged the firm understanding of the relationship between superfluidity and superconductivity that now exists, and which provides the foundation for investigating unconventional superconductors and superfluid phases.

Allen and Misener observed the flow of liquid helium-4 through long, thin tubes, and found that the fluid's viscosity became immeasurably low at temperatures below 2.17 kelvin. Kapitza obtained similar results by measuring the flow through a small gap between two glass disks (Fig. 1). With foresight, Kapitza noted a possible connection to superconductivity, for which a complete theory was eventually realized³ in 1957 by Bardeen, Cooper and Schrieffer (BCS). Shortly after the two *Nature* papers were published, an explanation for the superfluidity of liquid helium-4 was offered: Bose–Einstein condensation⁴, the process whereby many particles known as bosons ‘condense’ into a single quantum state.

In the quantum world, particles of the same type are indistinguishable, and there are only two classes of particle: fermions and bosons. However, an even number of interacting fermions can make a composite boson — for example, an atom of helium-4 is a composite boson that comprises six fermions (two protons, two neutrons and two electrons). At sufficiently low temperatures, helium-4

atoms undergo Bose–Einstein condensation and become a superfluid. Similarly, in the BCS theory of superconductivity, electrons that have a suitably attractive interaction can combine into charged composite bosons called Cooper pairs, which condense to form a superconductor.

In the wake of the Second World War, substantial quantities of the light isotope of helium, helium-3, became available through production of the heavy isotope of hydrogen (hydrogen-3 or tritium) for use in the hydrogen bomb. Because helium-3 contains an odd number of fermions (two protons, one neutron and two electrons), it is not a composite boson. It might therefore be considered that Bose–Einstein condensation could not take place and that helium-3 could never be a superfluid. However, the success of the BCS theory suggested another possibility: composite bosons comprising Cooper pairs of helium-3 atoms might condense into a superfluid, much like the electrons of a BCS superconductor.

The properties of this hypothetical superfluid were studied theoretically^{5–7} in the 1960s. Research on the subject then exploded following the unexpected discovery⁸ in 1972 of this superfluid at temperatures below 0.003 K. At first, the observations were interpreted as spontaneous nuclear magnetic ordering in solid helium-3, but shortly afterwards, they were correctly identified as the transition to a superfluid⁹. Nuclear magnetic ordering in solid helium-3 was discovered¹⁰ two years later at a temperature of 0.001 K.

Cooper pairs have two types of angular momentum, characterized by the orbital quantum number (L) and the spin quantum number (S). Conventional BCS superconductors have $L = 0$ and $S = 0$, whereas superfluid helium-3 has $L = 1$ and $S = 1$. Nevertheless, the superfluid's properties can be understood using a modified version of the BCS theory¹¹. The discovery of superfluid helium-3



50 Years Ago

There was an increase in the number of patients discharged from British hospitals in 1964, and a decrease in the average length of stay in hospital compared with 1962 and 1963. Men and boys stayed in hospital an average length of 18.3 days in 1964; women and girls ... averaged just under two days less (16.7 days) ... These are some of the findings in ... the *Report on Hospital In-Patient Enquiry* for the year 1964 ... The report contains detailed tables prepared from the 1964 ten per cent sample of discharges and deaths recorded ... The tables are a mine of information ... Injuries, poisonings and the like are all analysed in great detail according to whether they were caused by road traffic accidents, accidents in the home, or “other” mishaps.

From *Nature* 27 January 1968

100 Years Ago

It was stated officially ... that the Admiralty had tested many methods of disguising mercantile shipping. One of these methods is to paint the ship with various quaint combinations of different colours. But this does not appear to have proved much of a success ... Mr. Abbott H. Thayer ... was one of the first to recognise that a high degree of invisibility is conferred on certain birds by the simple adaptation of being dark above and whitish below. He took two wooden decoy ducks, and placed them against a sandbank. One was coloured like the sand ... the other was coloured on its upper parts darker than the surrounding sand, and graded below to pure white. At a short distance the first was still clearly visible, but the second was quite lost against its background ... Some modification of this experiment has been tried on ships ... but this device has not proved so successful as had been hoped.

From *Nature* 24 January 1918

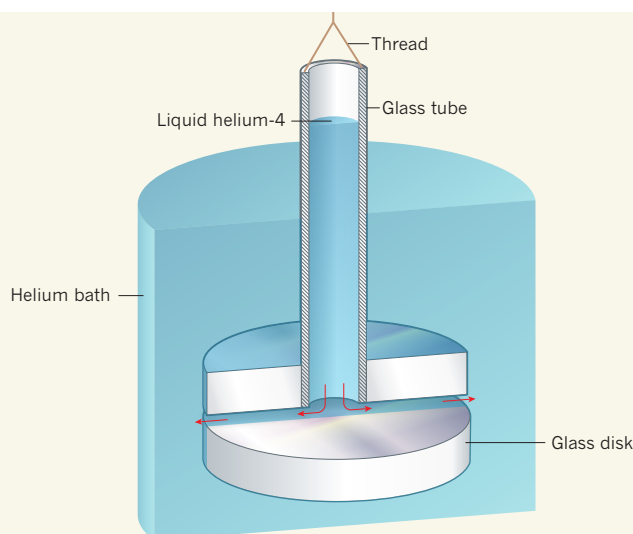


Figure 1 | Experimental evidence for superfluidity. In 1938, Allen and Misener¹ and Kapitza² showed that liquid helium-4 becomes a superfluid — a fluid with zero viscosity — at very low temperatures. Whereas Allen and Misener measured the flow of liquid helium-4 through long, thin tubes, Kapitza observed the flow (red arrows) from a glass tube to a helium bath, through a narrow gap between two glass disks. The separation between the disks was adjusted using a thread such that the level of the column of liquid in the glass tube was above the level of the helium bath. At temperatures above 2.17 kelvin, Kapitza found that the difference in height between these levels was maintained for several minutes. Conversely, at lower temperatures, the difference disappeared in seconds. Kapitza concluded that the viscosity of liquid helium-4 must be immeasurably low below 2.17 K. (Figure adapted from ref. 2.)

therefore marked the birth of unconventional superconductivity — and, more precisely, of superfluids that break certain fundamental symmetries of the normal (non-superfluid) state. The non-zero values of L and S in superfluid helium-3 correspond to broken rotational and time-reversal symmetries, which cause the substance to have a non-trivial topology.

In the absence of a magnetic field, superfluid helium-3 has two phases: A and B, with the B phase dominating the pressure–temperature phase diagram (a graph that plots the physical state of a material at various pressures and temperatures). The B phase can exist in many excited states, as a consequence of broken rotational symmetry associated with the total angular momentum of Cooper pairs^{7,12,13}. The states of the B phase are classified by total angular momentum quantum numbers (J) of 0, 1 and 2. The $J=2$ state comprises bosons that are analogous to the famous Higgs boson¹⁴. A remarkable finding is that the broken symmetry of the B phase, and its $J=2$ state, enable the propagation of transverse sound waves^{15,16} — a feature that was unheard of in liquids and was often assumed to be a property only of rigid solids.

Since the discovery of superfluid helium-3, many unconventional superconductors have been found. The best known are copper oxide compounds known as cuprates, which have the quantum numbers $L=2$ and $S=0$, and certain heavy fermion compounds¹⁷. However, only one superconducting compound, the uranium–platinum system UPt₃, has been

discovered that has more than one superfluid phase, like helium-3. UPt₃ has $L=3$ and $S=1$, as predicted¹⁸, and one of its phases breaks time-reversal symmetry in a similar way¹⁹ to the A phase of helium-3.

In the past few years, helium-3 has been shown to exhibit new superfluid phases when confined to low-density materials called aerogels, small pores and narrow slabs²⁰. Such phases are being investigated further. Eighty

years after the discovery of superfluidity in liquid helium-4, the search is on for other scientifically interesting superfluids and superconducting materials. ■

William P. Halperin is in the Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA. e-mail: w-halperin@northwestern.edu

1. Allen, J. F. & Misener, A. D. *Nature* **141**, 75 (1938).
2. Kapitza, P. *Nature* **141**, 74 (1938).
3. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. *Phys. Rev.* **108**, 1175–1204 (1957).
4. London, F. *Nature* **141**, 643–644 (1938).
5. Anderson, P. W. & Morel, P. *Phys. Rev.* **123**, 1911–1934 (1961).
6. Balian, R. & Werthamer, N. R. *Phys. Rev.* **131**, 1553–1564 (1963).
7. Vdovin, Y. A. in *Application of Methods of Quantum Field Theory to Problems of Many Particles* (ed. Alekseyeva, A. I.) (GOS ATOM ISDAT, 1963).
8. Osheroff, D. D., Richardson, R. C. & Lee, D. M. *Phys. Rev. Lett.* **28**, 885–888 (1972).
9. Osheroff, D. D., Gully, W. J., Richardson, R. C. & Lee, D. M. *Phys. Rev. Lett.* **29**, 920–923 (1972).
10. Halperin, W. P., Archie, C. N., Rasmussen, F. B., Buhrman, R. A. & Richardson, R. C. *Phys. Rev. Lett.* **32**, 927–930 (1974).
11. Vollhardt, D. & Wölfle, P. *The Superfluid Phases of Helium 3* (Taylor & Francis, 1990).
12. Maki, K. J. *Low Temp. Phys.* **16**, 465–477 (1974).
13. Serene, J. W. III *Theory of Collisionless Sound in Superfluid Helium 3*. PhD thesis, Cornell Univ. (1974).
14. Volovik, G. E. & Zubkov, M. A. *J. Low Temp. Phys.* **175**, 486–497 (2014).
15. Moores, G. F. & Sauls, J. A. *J. Low Temp. Phys.* **91**, 13–37 (1993).
16. Lee, Y., Haard, T. M., Halperin, W. P. & Sauls, J. A. *Nature* **400**, 431–433 (1999).
17. Norman, M. R. in *Novel Superfluids Vol. 2* (eds Bennemann, K.-H. & Ketterton, J. B.) Ch. 13 (Oxford Univ. Press, 2014).
18. Sauls, J. A. *Adv. Phys.* **43**, 113–141 (1994).
19. Schemm, E. R., Gannon, W. J., Wishne, C. M., Halperin, W. P. & Kapitlnik, A. *Science* **345**, 190–193 (2014).
20. Lee, Y. & Halperin, W. P. *J. Low Temp. Phys.* **189**, 1–14 (2017).

This article was published online on 15 January 2018.

BIOTECHNOLOGY

Kiss-and-tell way to track cell contacts

Transient cellular contacts are essential for the generation of an immune response, but these are difficult to measure *in vivo*. A labelling technique now offers a way to record such interactions between cells. SEE LETTER P.496

AARON P. ESSER-KAHN

Contact between two cells is a key step in the transfer of information during biological processes. However, monitoring dynamic cellular interactions *in vivo* poses many technical challenges. On page 496, Pasqual *et al.*¹ report the development of a technique that can track interactions between

cells that contact each other through receptor–ligand binding.

A key step in the development of an immune response involves contact between an antigen-presenting cell (APC), such as a dendritic cell, and an immune cell called a T cell. On the APC surface, a receptor called the major histocompatibility complex (MHC) displays a protein fragment known as an antigen. If the

antigen is recognized by the receptor on the T cell, an immune response is triggered. Such cellular interaction is essential for the success of vaccination, cancer immunotherapy and the elimination of disease.

Pasqual and colleagues now describe a method that can quantify the interactions between APCs and T cells *in vivo*. The ability to count the frequency and number of interactions is fundamental to analysing many complex networks. For networks as diverse as Facebook, academic citations and molecular interactions in a biochemical pathway, such measurements are the main way of assessing the importance of an interaction. And yet for the immune system, which is key to good health, a simple tool to allow this has been lacking.

One current approach for mapping cellular interactions involves an enzyme-based labelling technique that measures static connections between neuronal cells grown in culture². The authors describe an advance on this approach, using a form of enzyme-facilitated interaction mapping that is suited to the transient cellular interactions found in the immune system. Their method tracks the enzymatic transfer of a molecular label containing a small amino-acid tag, attached to an easily monitored molecule such as biotin. The molecular label can be transferred from one cell to another only if the cells are close enough together for an interaction to occur between a receptor and a ligand on the surfaces of the interacting cells. The molecular tag can then be detected by standard cell-analysis methods such as microscopy, or quantified *in vitro* using fluorescence analysis — tools already available to most biological researchers. The authors refer to this method of tracking a molecular 'kiss' between cells as LIPSTIC.

As a testing ground for their approach, Pasqual and colleagues choose a key interaction on the surface of immune cells that is highly dynamic and yet not physically involved in antigen presentation: the contact between a CD40L ligand, which is present on T cells, and its binding partner, the CD40 receptor, on APCs. Using mice, the authors engineered a fusion protein containing a sortase enzyme and CD40L, and generated a version of CD40 that contained glycine amino-acid residues at its amino terminus. The sortase was supplied with a labelling tag that became bound to the enzyme. In this system, when CD40 and CD40L interact, the sortase on the T cell attaches the tag to an N-terminal glycine residue on the APC's CD40 (Fig. 1).

LIPSTIC offers three major advances for the field. First, it allows the level of cell–cell interactions in the immune system to be quantified — the more APCs and T cells that interact, the higher the amount of cell labelling that is detected. Therefore, LIPSTIC provides a direct measure of a key step in the initiation

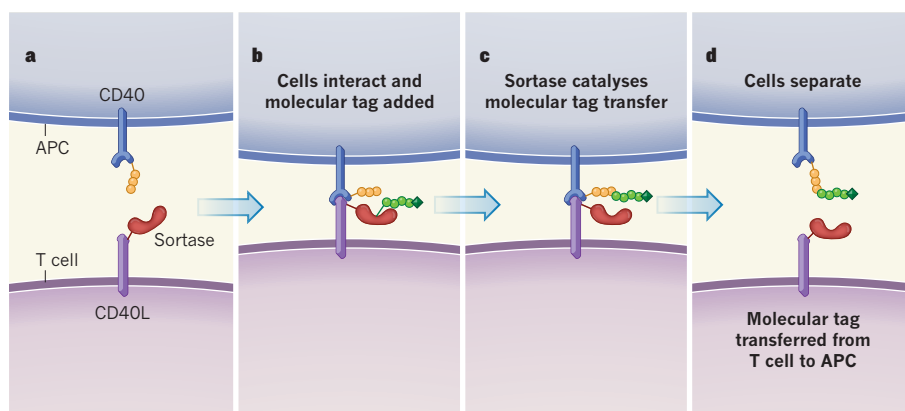


Figure 1 | Tracking cellular interactions. Pasqual *et al.*¹ describe a technique (termed LIPSTIC) that can monitor interactions between a T cell and an antigen-presenting cell (APC). **a**, Using mice for *in vivo* experiments, the authors generated T cells containing the CD40L ligand fused to the enzyme sortase, and APCs in which the CD40 receptor contains a few glycine amino-acid residues (yellow) at its amino terminus. **b**, When a CD40–CD40L interaction occurs between the cells, if a molecular tag consisting of a few amino acids (green circles) and the molecule biotin (green square) is added, the tag attaches to sortase. **c**, The enzyme then catalyses the transfer of the tag to a glycine on the amino terminus of CD40. **d**, When the cells separate, their interaction can be tracked by the presence of the transferred tag.

of an adaptive immune response. It improves on current methods that measure this step indirectly, such as monitoring of the levels of inflammatory cytokine proteins or antibody production, which assess only the downstream effects of such interactions.

Second, LIPSTIC might offer the possibility of identifying the types of T cell with which APCs interact. Interactions between APCs and different T-cell types can determine both the nature and magnitude of an immune response. For example, the degree of activation of T cells that express the protein CD8 can provide a way of assessing the effectiveness of cancer immunotherapy³. Improved understanding of the interactions between T cells and APCs might thus allow the development of more-effective cancer immunotherapies and vaccines.

Third, and perhaps most impressively, the necessary tools and instrumentation for LIPSTIC analysis are readily accessible. This approach could therefore be rapidly implemented without the technology-transfer delays that often slow the adoption of a technical innovation.

To test LIPSTIC's usefulness for providing biological insights into immune-system function, the authors analysed the APC–T-cell interactions. Surprisingly, they found that these cells have two modes of interaction, although it had been thought that interaction occurs only when an antigen-bound MHC is presented to the T-cell receptor. The authors observed interaction between APCs and T cells that did not require an antigen-loaded MHC; the label was transferred onto cells that were not loaded with antigen. This previously unknown interaction would be difficult to observe without a method such as LIPSTIC. Why does it occur, and what purpose does it serve? The answers could have implications for

efforts to improve immune responses.

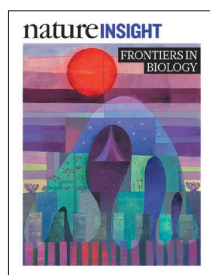
Despite LIPSTIC's evident potential, many challenges remain that will determine the impact of this technique on the wider field of study of cell–cell interactions. How well will it work if adapted for use in systems other than those tested by Pasqual and colleagues? Another challenge will be to determine the level of nonspecific background labelling and of labelling errors inherent in the LIPSTIC approach. The ability to assess both the accuracy and precision of a labelling method is needed for all good quantitative tools.

One way in which the authors have already started to address the specificity of labelling is by using a sortase that has a low affinity for N-terminal glycines. However, the body is full of compounds that are similar to N-terminal glycines. Sortase, although specific for protein labelling *in vitro*, has not previously been used as a labelling tool in as reactive or demanding an environment as a whole organism. The potential for sortase to transfer a labelling tag to other biological entities at a low background level should be examined more fully. Nevertheless, the hitherto secret world of interactions between T cells and APCs can now be dissected and studied. LIPSTIC offers a way of quantifying contact, one of the most mysterious, but key, elements of a cellular network. ■

Aaron P. Esser-Kahn is at the Institute for Molecular Engineering, University of Chicago, Chicago, Illinois 60615, USA.
e-mail: aesserkahn@uchicago.edu

1. Pasqual, G. *et al.* *Nature* **553**, 496–500 (2018).
2. Liu, D. S., Loh, K. H., Lam, S. S., White, K. A. & Ting, A. Y. *PLoS ONE* **8**, e52823 (2013).
3. Mellman, I., Coukos, G. & Dranoff, G. *Nature* **480**, 480–489 (2011).

This article was published online on 17 January 2018.



Cover illustration

'Differentiation landscape' by Nik Spencer

Editor, *Nature*

Philip Campbell

Publishing

Richard Hughes

Insights Editor

Ursula Weiss

Subeditors

Rachel Jones,
Nicola Bailey

Art Editor

Nik Spencer

Sponsorship

Reya Silao

Production

Ian Pope

Marketing

Steven Hurst

Editorial Assistant

Mary Craig

The Campus
4 Crinan Street
London N1 9XW, UK
Tel: +44 (0) 20 7833 4000
e: nature@nature.com

**SPRINGER
NATURE**

The *Nature* Insight 'Frontiers in Biology' aims to cover timely and important developments across biology, ranging from subcellular molecular mechanisms to whole-organism physiology and biomedicine.

The development of mature blood cells from haematopoietic stem cells (HSCs) has long served as a paradigm for stem-cell research, with the haematopoietic differentiation tree being widely used as a model for the maintenance of hierarchically organized tissues. In this first Review, Laurenti and Göttgens discuss how recent results and new technologies have challenged some aspects of our classical view of blood production from HSCs, and enhanced our understanding of the mechanisms by which stem and multiprogenitor cells contribute to mature blood cell types.

Our skin serves as the interface between us and the environment, and is inhabited by a complex microbiota whose interactions with the host range from commensal or mutualistic to pathogenic. Chen, Fischbach and Belkaid discuss how the interactions between skin-resident microbes and their hosts are shaped by the specific context in which they take place. Factors such as the immune status of the host, tissue localization, and microbe-microbe interactions determine whether the consequences are beneficial or detrimental to the host, contributing to health or disease.

Organoids are self-assembling 3D tissue cultures that can capture normal and abnormal organogenesis *in vitro*. In this third Review, Sergiu Pașca surveys the most recent advances and remaining challenges in these technologies, with emphasis on brain development and the generation of complex neural circuits.

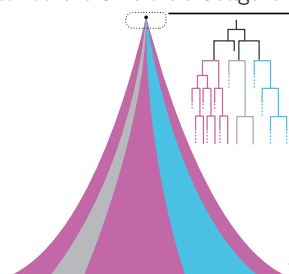
In the final Review of this collection, Herbst, Morgensztern and Boshoff review the aetiology, key biological features, and major therapy approaches for lung cancer over the past two decades, with a particular focus on recent developments of molecularly targeted and immunotherapies.

Nathalie Le Bot, Barbara Marte & Ursula Weiss
Senior Editors

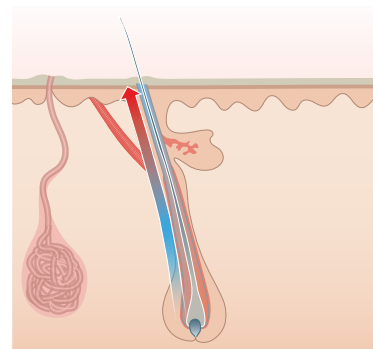
CONTENTS

REVIEWS

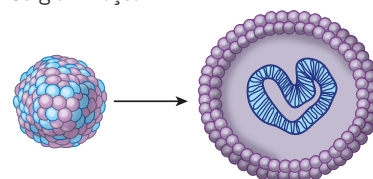
- 418 From haematopoietic stem cells to complex differentiation landscapes**
Elisa Laurenti & Berthold Göttgens



- 427 Skin microbiota-host interactions**
Y Erin Chen, Michael A Fischbach & Yasmine Belkaid



- 437 The rise of three-dimensional human brain cultures**
Sergiu P Pașca



- 446 The biology and management of non-small cell lung cancer**
Roy S Herbst, Daniel Morgensztern & Chris Boshoff

From haematopoietic stem cells to complex differentiation landscapes

Elisa Laurenti¹ & Berthold Göttgens¹

The development of mature blood cells from haematopoietic stem cells has long served as a model for stem-cell research, with the haematopoietic differentiation tree being widely used as a model for the maintenance of hierarchically organized tissues. Recent results and new technologies have challenged the demarcations between stem and progenitor cell populations, the timing of cell-fate choices and the contribution of stem and multipotent progenitor cells to the maintenance of steady-state blood production. These evolving views of haematopoiesis have broad implications for our understanding of the functions of adult stem cells, as well as the development of new therapies for malignant and non-malignant haematopoietic diseases.

When Ernst Haeckel first used the word stem cell ('Stammzelle') in 1868, as a Darwinist he used it to refer to the primordial unicellular organism from which all multicellular life descended¹. This stem cell therefore sat at the root of a branching family tree, incidentally called a stem tree in German ('Stammbaum', meaning a tree that shows where things stem from). Shortly thereafter, Haeckel's biogenetic law, in which ontogeny recapitulates phylogeny, prompted him to use the stem-cell term to describe the fertilized egg. Histopathologists subsequently applied this stem-cell concept to normal and leukaemic haematopoiesis, putting forward the concept of a common progenitor of red and white blood cells² as well as a common precursor of myeloid and lymphoid leukaemic cells³. From the very beginning, the stem-cell concept has thus been framed into a tree-like model, in which multipotent stem cells give rise to their progeny through an ordered series of branching steps.

The first *in vivo* assay for stem-cell function was based on the rescue of lethal irradiation by bone marrow transplantation⁴, followed by the first estimation of stem-cell numbers by counting haematopoietic colonies in the spleens of transplanted mice. This not only provided an estimate of the frequency of spleen colony-forming units at 1 in 10,000 bone marrow cells⁵, but also delivered the first definitive proof for *in vivo* multipotent progenitor cell function based on tracking cytogenetic abnormalities within individual spleen colony-forming units⁶. Fluorescence-activated cell sorting subsequently facilitated the purification of transplantable haematopoietic stem cells (HSCs), with a landmark 1988 publication⁷ that demonstrated the use of positive and negative selection. HSCs have historically been defined on the basis of two essential properties: self-renewal and multipotency. Operationally, this is tested via transplantation experiments. By contrast, progenitors are defined by the absence of extended self-renewal and a restricted lineage differentiation capacity (most often bi- or unilineage), so that they are usually lost within the first 2–3 weeks after transplantation⁸.

Around the year 2000, the characterization of progenitor populations downstream of HSCs resulted in a model of the haematopoietic differentiation tree that is still shown in many textbooks today (Fig. 1a). In this model, the first branch point segregates lymphoid potential from all other lineages (myeloid, erythroid and megakaryocytic), followed by several further branching steps on either side of the tree progressing from multi- to bi- and finally to unipotent progenitor cells. The subsequent introduction of other surface markers suggested several modifications of this classical tree, including lymphoid and myeloid fates

remaining associated until further down the tree^{8–10}, early megakaryocyte branching^{11,12} as well as subdivision of the multipotent progenitor compartment into distinct subpopulations^{13,14} (Fig. 1b). Moreover, the picture is further complicated because the HSC pool itself is functionally and molecularly heterogeneous^{11,12,15–20}. These studies are most advanced in the mouse system, in which we now have what may seem a bewildering number of different structures for the haematopoietic tree. Although it is likely that all these structures capture true aspects of HSC differentiation, collectively they would be difficult to squeeze into a single, rigid branching tree. New ways of not only thinking about, but also graphically representing the process of HSC differentiation are thus required. In this Review, we illustrate how new technologies are challenging the classical view of the haematopoietic hierarchy as a highly compartmentalised and stable structure. The emerging picture is one of a collection of heterogeneous populations organized hierarchically, with gradual progression from one to the next, and which remains highly flexible to meet the changing needs of blood demand.

Stem cell and progenitor boundaries

With self-renewal and multipotency at the heart of what defines an HSC, much research has been invested into understanding the underlying cellular and molecular processes.

Defining the HSC state

At the cellular level, switching off self-renewal coincides with the turning on of lineage programs. It thus seemed plausible that this would also be true at the molecular level, and the concept of multilineage priming was proposed early on as a possible underlying molecular mechanism by which HSCs maintain multipotentiality²¹ (Box 1). However, the advent of genomic technologies^{14,22–25} coupled with mouse genetic studies has demonstrated that the HSC transcriptional programme is defined by a collection of unique metabolic and cellular properties, which are not intuitively linked directly with multipotency. Approximately 70% of all expression changes between HSCs and early progenitors occur independently of lineage choice²⁴, with a similar dichotomy also at the levels of methylation^{26–29} and chromatin accessibility³⁰. Correspondingly, HSCs reside in a quiescent^{19,20,31,32}, autophagy-dependent^{33,34} and glycolytic^{35,36} state marked by low mitochondrial activity^{37,38} and tightly controlled levels of protein synthesis, below those of most other haematopoietic cell types³⁹. Stem-cell-specific stress response and quality-control mechanisms allow the preservation of the integrity of the HSC compartment

¹Department of Haematology and Wellcome and MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK.

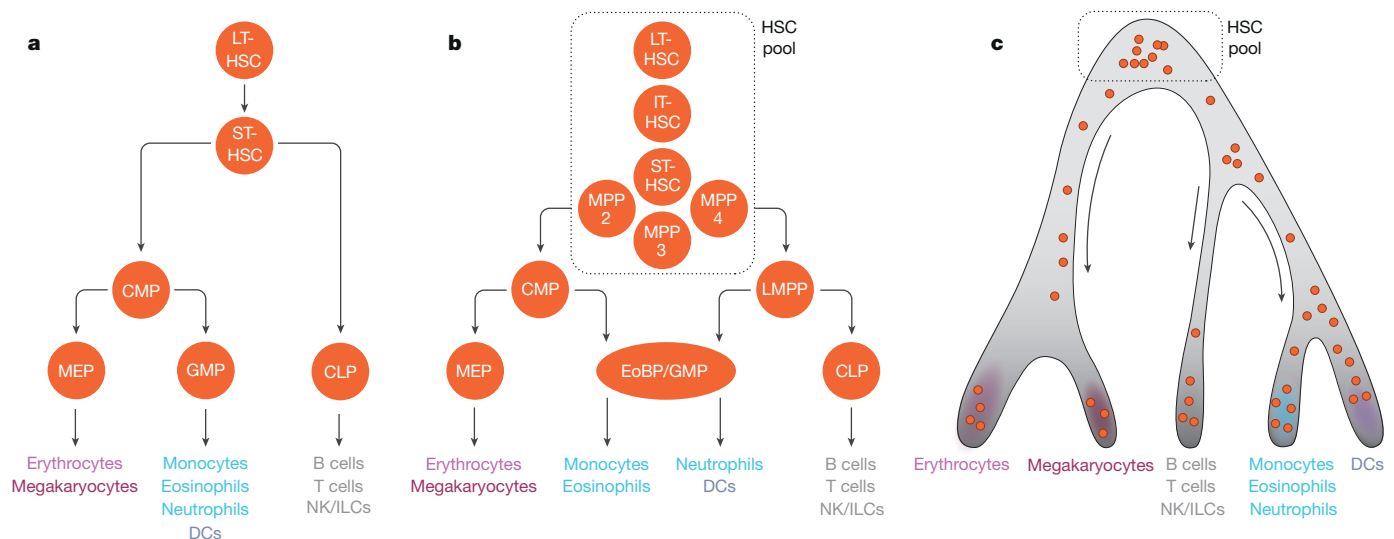


Figure 1 | Timeline of hierarchical models of haematopoiesis.

a, Visualization based on cutting-edge research around the year 2000: HSCs are represented as a homogeneous population, downstream of which the first lineage bifurcation separates the myeloid and lymphoid branches via the common myeloid progenitor (CMP) and common lymphoid progenitor (CLP) populations. DCs, dendritic cells; EoBP, eosinophil–basophil progenitor; GMP, granulocyte–monocyte progenitors; LT, long-term; ILCs, innate lymphoid cells; MEP, megakaryocyte–erythrocyte progenitors; NK, natural killer cells; ST, short-term. **b**, During the years

2005–2015, this visualization incorporates new findings: the HSC pool is now accepted to be more heterogeneous both in terms of self-renewal (vertical axis) and differentiation properties (horizontal axis), the myeloid and lymphoid branches remain associated further down in the hierarchy via the lymphoid-primed multipotential progenitor (LMPP) population, the GMP compartment is shown to be fairly heterogeneous¹⁴². **c**, From 2016 onwards, single-cell transcriptomic snapshots indicate a continuum of differentiation. Each red dot represents a single cell and its localization along a differentiation trajectory.

after exposure to DNA^{40,41} and protein damage^{42,43} or metabolic stress^{33,34}. By contrast, progenitors are highly proliferative and metabolically active cells that are dependent on oxidative metabolism and mitochondrial function.

Importantly, so-called HSC-specific characteristics are not absolute: HSCs occasionally divide during homeostasis and get activated in response to stress, and therefore transiently pass through a proliferative state. It is also worth noting that early lymphoid progenitors (mouse lymphoid-primed multipotential progenitors⁹ or human myeloid-lymphoid progenitors⁸) share common transcription networks with HSCs²⁴, which in the former push towards B cell differentiation²⁴ while in the latter inhibit self-renewal^{44,45}. Although the induction of lineage-specific transcriptional programs may occur largely independently of the loss of stem-cell characteristics, regulators such as RUNX1 can be involved in both⁴⁶, suggesting that much remains to be learned about how lineage decisions are coordinated with changes of cellular state.

HSC self-renewal and cell cycle interplay

Whether the variability in HSC outputs is an intrinsic system property, a reflection of stochastic behaviour, environmental influences, technically related variability or a combination of all of these, has been and still is a subject of debate. Nonetheless, in the past 10–15 years, this heterogeneity in behaviour has been notably formalized, and it is now accepted that HSCs are heterogeneous in terms of durability of engraftment upon transplantation, cell cycle properties and differentiation (Table 1).

HSCs first of all differ in the degree to which they self-renew, equating to the number of symmetric divisions that an HSC can make over its lifetime. Operationally, the field now widely accepts that, both in mouse and human, HSCs that repopulate in transplantation assays for more than 16 weeks in a primary transplantation and at least in a second round of transplantation are considered to be long-term HSCs^{17,47–50}. If cells can produce all differentiated cell types and engraft transiently in primary (and in some cases secondary) transplants, they are referred to as intermediate HSCs¹⁶, short-term HSCs or multipotent progenitors (MPPs)^{13,14,47}, depending on the length and robustness of the graft produced.

Heterogeneity in self-renewal capacity seems to be directly correlated to the time HSCs spend in quiescence. Label-retaining studies have demonstrated that the most dormant HSCs (which retain the label over

months at the steady-state) display the most robust and longest repopulation capacity^{19,20,51–53}. Two cell cycle parameters are inversely correlated with repopulation capacity: the frequency of division (length of interval between divisions), but also the time a single HSC takes to exit quiescence *in vitro*^{16,54}. Of note, the level of pre-existent *CDK6* mRNA and *CDK6* protein in quiescent HSCs directly determines the kinetics of quiescence exit⁵⁴, and may thus serve as a marker of the quiescent state^{32,54} (less *CDK6* corresponds to higher dormancy). Dormancy is also associated with high levels of vitamin A metabolism via retinoic acid signalling³², high levels of p57^{32,55}, low levels of protein synthesis³², low MYC activity^{32,56}, and may exist as a continuum of quiescent states between the most dormant HSCs and their activated counterparts³². Although dormant HSCs forced into activation by stress signals can return to dormancy^{19,51}, Bernitz *et al.* estimated that four divisions in adulthood are sufficient for irreversible loss of self-renewal⁵³.

Several label-retaining assays have been developed to allow the isolation of HSCs based on their division history^{19,20,32,51–53,57}. A major limitation of all label-retaining studies so far is that the rate of symmetric divisions is inferred or indirectly measured at the bulk level. However, direct experimental measurement of asymmetric versus symmetric label-diluting divisions will be required to understand the dynamics at play within the HSC compartment. New integrative tools, most likely at the single-cell level, will have to be developed to address this challenge. Nonetheless, it is already apparent that distinct cell cycle properties within the HSC pool are intimately linked to HSC function. Coupled with a wide range in division frequencies (depending on the HSC subset, from once a month to twice a year in mouse^{19,51,53,58}), this means that there will be substantial variation in the contribution of distinct HSC subsets to blood formation.

Heterogeneity in HSC lineage output

The capacity to give rise to all differentiated blood cell types is a fundamental aspect of what constitutes an HSC. It is, however, now accepted that there are distinct differentiation behaviours within the HSC and MPP^{13,14} compartments (Table 1, Fig. 1b). Using limiting-dilution analysis and single-cell transplantation, the Müller–Sieburg and Eaves groups described HSCs that differ in their relative myeloid and lymphoid

BOX 1

Mechanistic underpinnings of cell-fate choice

Cell-fate decisions entail a choice between alternative gene expression programs, commonly executed by transcriptional and epigenetic regulators in response to extracellular signals. Transcription factors bind to specific sequence motifs within promoter, enhancer and silencer regions. Cell-type-specific expression is achieved through combinatorial transcription factor interactions, which form key building blocks of wider regulatory networks. Transcription factor complexes recruit epigenetic regulators to modulate the activation status of a gene locus, which can be transmitted to subsequent cell generations as 'epigenetic memory'. Importantly, a progenitor cell can become primed by opening up regulatory elements associated with genes that drive differentiation down a specific mature lineage¹⁴⁴.

Multipotent cells are thought to exhibit multilineage priming, which entails simultaneous, low-level activation of expression programs for alternative lineages. Lineage choice then constitutes one program 'winning out' while the alternative program is extinguished. Cross-antagonism between pairs of lineage-determining transcription factors represents an attractive mechanistic model, which initially focused on the erythroid regulator GATA1 and the myeloid regulator PU.1. More recent single-cell time-lapse imaging, however, questioned whether an erythroid/myeloid fate choice is indeed driven primarily by cross-antagonism between GATA1 and PU.1¹⁴⁵. Evidence for lineage priming and its resolution by cross-antagonism has been reported at single-cell resolution for other transcription factor pairs, such as the neutrophil or macrophage fate choice that is controlled by GFI1 and IRF8⁸⁹.

Both instructive and stochastic models have been proposed as mechanisms that trigger the upregulation of one lineage program over another. Stochastic here generally refers to random, unequal distribution of molecules after cell division. Instructive models posit that low-level expression of a cytokine receptor is sufficient for a cell to be responsive to external signals, as shown for several myeloid cytokines¹⁴⁶. Autoregulation and feed-forward loops also have important roles in cell-fate choice decisions. For example, positive autoregulation of GATA1 stabilizes erythroid fate¹⁴⁷, positive autoregulation of PU.1 stabilizes myeloid identity¹⁴⁸, and the highly connected triad of GATA2, TAL1 (also known as SCL) and FLI1 is thought to stabilize the stem/progenitor state^{149,150}. In feed-forward loops, an upstream regulator induces its target directly as well as through an intermediate regulator¹⁵¹. Feed-forward motifs can filter out transient signals, and when coupled with autoregulation, can generate forward momentum.

output^{15,59}. More recently, the Jacobsen and Nakauchi groups identified HSCs that predominantly differentiate towards megakaryocytes and platelets (platelet-biased)^{11,12}. Subsequent transplantation experiments have highlighted that single HSCs execute only a limited repertoire of lineage fates patterns, and the only long-term unilineage read-out observed in this study was that of platelets⁶⁰. There are limitations to single-cell transplantation, particularly as very low contributions to certain lineages may be missed. Moreover, a cell with unilineage read-out may have the potential to give rise to other lineages in a durable fashion in other conditions. HSCs with platelet-biased output in unperturbed haematopoiesis acquired broader potential after transplantation⁶¹, and similarly HSCs with platelet-biased output after transplantation could produce myeloid and lymphoid cells *in vitro*⁶⁰. Finally, both platelet-biased HSCs and long-lived platelet progenitors^{11,62,63} may coexist and be difficult to distinguish.

Importantly, HSC heterogeneity is not simply stochastic, as it can be propagated by serial transplantation^{15,18,64,65}, indicating intrinsic programming, the molecular basis for which remains unclear. As discussed later, these findings have important conceptual implications, because they question at what cellular stage lineage choices occur. Recent evidence

suggests that the overall picture may be even more complex, with distinct metabolic needs⁶⁶ and clonal expansion capacity of HSCs. There are examples in which HSCs with high clonal expansion capacity generate subsets with lower output⁶⁷, but also cases in which HSCs with very modest clonal expansion in the first transplantation, generate the most robust grafts upon serial transplantation^{68,69}.

Rethinking blood lineage relationships

Several recent studies at the single-cell level have questioned the routes by which lineage differentiation occurs.

Single-cell assays to study potential

Cell-fate decisions (Box 1) are executed at the level of individual single cells. To understand their regulation, it is therefore imperative that both the biological assays testing their cellular function as well as the biochemical assays examining their molecular profiles are performed at single-cell resolution. Although most advanced in the mouse, *in vivo* transplantation assays have been and remain fundamental for our understanding of HSC biology, as the only assay that can test for HSC self-renewal. Because of the suboptimal support of human cells from the mouse microenvironment, xenotransplants cannot robustly read-out all possible differentiation routes, particularly not at the single-cell level. The last 10 years have therefore seen a collective effort in defining the lineage potential of single human progenitor cells by using highly defined *in vitro* models that can support the differentiation of most mature blood cell types. Work from the Dick and Vyas groups, together with studies in mice from the Jacobsen group, suggested that the first restriction in lineage potential does not segregate lymphoid and myeloid potential as postulated by the common myeloid and lymphoid progenitor models (Fig. 1a), but that these potentials remain coupled in the lymphoid-primed multipotential progenitor^{9,10} and myeloid-lymphoid progenitor⁸ compartments (Fig. 1b). Many other subpopulations, most with either bi- or unilineage capacity, have since been found in both the lympho-myeloid branch⁷⁰ and the myelo-erythroid-megakaryocytic branch⁷¹⁻⁷⁵. Altogether, it seems that few single cells read-out as multipotential within the progenitor compartment. There are, however, some caveats with such interpretations, because strong instructive signals provided by the *in vitro* cultures, or potentially high stress levels that HSCs are exposed to during single-cell transplants may promote unilineage output. Moreover, a given cell may be bipotential based on its molecular state, but if it makes a lineage choice before dividing, it will read-out as unipotent in functional assays. The evidence so far suggests that lineage choice occurs earlier than previously thought, and as recently shown for dendritic cells⁷⁶, probably as early as within the phenotypic HSC populations.

Single-cell transcriptional landscapes

Single-cell expression analysis of RNA was first reported over 25 years ago⁷⁷, but remained low throughput in the number of genes and cells until microfluidic approaches were introduced. These quickly prompted studies that reported the expression of dozens of genes in hundreds of single HSCs and haematopoietic stem and progenitor cells (HSPCs)⁷⁸, and provided new insights into core regulatory circuits⁷⁹, new progenitor populations^{72,73}, cellular hierarchies in normal and transformed haematopoiesis⁸⁰, dissociation between self-renewal potential and activation of lineage programs⁸¹, and the molecular overlap between HSC populations purified with four different cell-sorting strategies⁸². It is not practical to assay more than 200 genes per single cell with PCR-based methods, and these genes need to be predefined, thus limiting the scope for new discoveries. A real breakthrough was therefore provided by technical innovations, which now make it possible to perform transcriptome-wide RNA sequencing (RNA-seq) in thousands of single cells.

Following a landmark paper reporting the transcriptomes for more than 2,600 mouse single myelo-erythroid progenitor cells⁸³, a subsequent report of 1,600 transcriptomes ranging from true long-term HSCs to progenitors of all major lineages focused on generic stem-cell functions such as metabolic and cell cycle status⁸⁴. Several algorithms have been

Table 1 | Phenotypic or functionally defined HSC subsets

Species	Name	Cell-surface phenotype	Self-renewal	Cell cycle properties	Differentiation	References
Mouse	LT-HSC	Lin [−] Sca1 ⁺ cKit ⁺ CD34 [−] CD150 ⁺ CD135 [−] CD48 [−] ± EPCR ⁺ ± Rho ^{lo}	High			17, 47, 48, 141
	IT-HSC	Lin [−] Sca1 ⁺ cKit ⁺ CD34 ^{lo} CD135 [−] Rho ^{lo} CD49b ^{hi}	Intermediate	Short G ₀ exit		16
	ST-HSC/MPP1	Lin [−] Sca1 ⁺ cKit ⁺ CD135 [−] CD150 [−] CD48 [−]	Low			
	α	NA	High	NA	Ly-deficient	15, 108
	β	NA	High	NA	Balanced	
	γ	NA	Intermediate	NA	My-deficient	
	δ	NA	Low	NA	My-deficient	
	MPP2	Lin [−] Sca1 ⁺ cKit ⁺ CD135 [−] CD150 ⁺ CD48 ⁺	Low	Similar	Ly-deficient	13, 14
	MPP3	Lin [−] Sca1 ⁺ cKit ⁺ CD135 [−] CD150 [−] CD48 ⁺	Low		Balanced	
	MPP4	Lin [−] Sca1 ⁺ cKit ⁺ CD135 ⁺ CD150 [−] CD48 ⁺	Low		Ly-biased	
	d-HSC	NA	Higher	Dormant		19, 20, 51–53
	a-HSC	NA	Lower	Activated		
Human	LT-HSC	Lin [−] CD34 ⁺ CD38 [−] CD45RA [−] CD49f ⁺ CD90 ⁺ ± Rho ^{lo}	High	Long G ₀ exit		50, 54
	ST-HSC/MPP	Lin [−] CD34 ⁺ CD38 [−] CD45RA [−] CD49f ⁺ CD90 ⁺	Low	Short G ₀ exit		
	CD34 [−] LT-HSC	Lin [−] CD34 [−] CD38 [−] CD93 ^{hi}	High	Highly quiescent		68

Listed are subtypes of HSCs defined based on combinations of cell-surface markers and/or function. A global interpretation of the literature on HSC biology is complicated by the fact that each study usually uses only one of the classification schemes, so the extent of overlap between HSC subsets remains to be clarified. This list is not exhaustive. EPCR, endothelial protein C receptor; IT, intermediate-term; LT, long-term; Ly, lymphoid; My, myeloid; NA, not assessed; ST, short-term.

developed^{85–89} based on the idea that single-cell transcriptomes represent snapshots of single cells as they traverse differentiation landscapes (Fig. 1c). A recent study on human bone marrow haematopoiesis comprehensively sampled the HSPC compartment⁹⁰. Computational predictions suggestive of early lineage restriction were underpinned by *in vitro* single-cell culture assays, which led the authors to propose a model in which acquisition of lineage-specific fates is a continuous process, and unilineage-restricted cells emerge directly from a continuum of low-primed undifferentiated HSPCs, without any major transition through the multi- and bipotent stages. This is supported by other studies^{71,83,87}, which highlighted the abundance of unipotent progenitors within compartments that, at the population level, are multipotent.

There are, however, caveats with this new model and its heavy reliance on single-cell RNA-seq (scRNA-seq). A major conundrum here is that combinations of surface markers can readily split the HSPC compartment into functionally distinct subpopulations, including within the space proposed to be a continuum of low-primed cells when analysed by scRNA-seq. One possible explanation is that there is a decoupling between steady-state mRNA and protein expression levels. The counter argument is that multiomics analysis of highly purified bulk HSPC populations has shown good concordance between mRNA and protein for most genes¹⁴. It is possible also that functional heterogeneity of HSCs is primarily determined at the epigenetic level. Future measurements of chromatin accessibility and histone marks ideally at single-cell resolution may reveal such mechanisms. A third possible explanation is that even though purifying cells based on protein markers followed by functional assays suggests clean splits into distinct cell types with different biological functions, the changes in biological functionality are in reality much more gradual, where there is no binary biological difference between, for example, long-term HSCs and MPPs, but instead any individual cell sits somewhere along a continuous spectrum. A fourth possible explanation is that current scRNA-seq data analysis methods are not effective at distinguishing between closely related cell types, because many of the shared biological processes (such as the cell cycle, metabolism, motility) generate substantial heterogeneity, which may exceed the number of differentially expressed genes between closely related stages of haematopoietic maturation. The degree to which early haematopoiesis is characterized by a continuum versus distinct populations therefore remains a question that requires further investigation.

New representations of haematopoiesis

An immediate challenge for the field is how all the recent findings can be reconciled into new graphical models that describe the hierarchical organization of haematopoiesis. It seems clear that a tree in which a circle depicts each successive restriction in potential and each circle is connected to a few others by arrows is an over-simplification. First, the circle is intuitively viewed as a homogeneous set of cells with specific characteristics,

a vision incompatible with the large degree of heterogeneity observed experimentally. Second, these trees indicate a restricted set of possible transitions between circles, which probably underestimates the possible differentiation journeys *in vivo*. Because cell-surface markers can highly enrich for particular behaviours (differentiation, self-renewal or proliferative output), a model in which all lineages branch out directly from the HSC compartment also seems unrealistic (Fig. 1c). We thus propose an alternative visualization (Fig. 2a), in which the trajectories of differentiation are mapped over the areas that have long been represented as circles, highlighting both the diversity in possible routes and the prevalence of early lineage choice. In addition, it is important to remember that one HSC will produce a very large number of progeny, which increases exponentially with each division, an element so far ignored in graphical representations of haematopoiesis (including Fig. 2a). Divisional histories are difficult to measure, and are likely to be heterogeneous, but should nevertheless be incorporated in future experimental and computational analyses, which could result in new graphical representations of the blood system (Fig. 2b).

Making blood at steady state and under stress

In addition to defining the routes of lineage differentiation, another important question is to understand the quantitative contributions of HSCs and progenitors to daily and emergency haematopoiesis.

Studying unperturbed haematopoiesis

Although the haematopoietic differentiation tree is widely used as a model for how a hierarchically organized tissue is maintained, it is important to remember that this tree was largely derived from experiments that measure cell potential in colony or transplantation assays, rather than cell fate during steady-state differentiation. However, just because a single cell gives rise to two lineages in a colony assay, this does not prove that the same cell, when left alone in an unperturbed bone marrow environment, would have done the same. To understand the dynamics of blood formation, cells can be individually tagged (for example, by retro- or lentiviral insertions or barcodes) and transplanted into recipient mice to measure the contribution of each clone over time. Progressively more sensitive methods have been used in mice, primates and humans^{91–95}, collectively supporting the lineage biases and/or restrictions observed in single-cell transplants. Importantly, only a limited number of HSCs produce the vast majority of the differentiated cells in a transplantation setting, consistent with studies of the divisional history of HSCs^{19,20,51–53,57}, which demonstrate that only the very rare most dormant HSC can provide life-long reconstitution after transplantation.

Substantial excitement was raised by new technologies to assess the lineage output of individual stem and progenitor cells in unperturbed haematopoiesis⁹⁶. Doxycycline-induced mobilization of a sleeping beauty transposon in stem and progenitor cells was used to generate

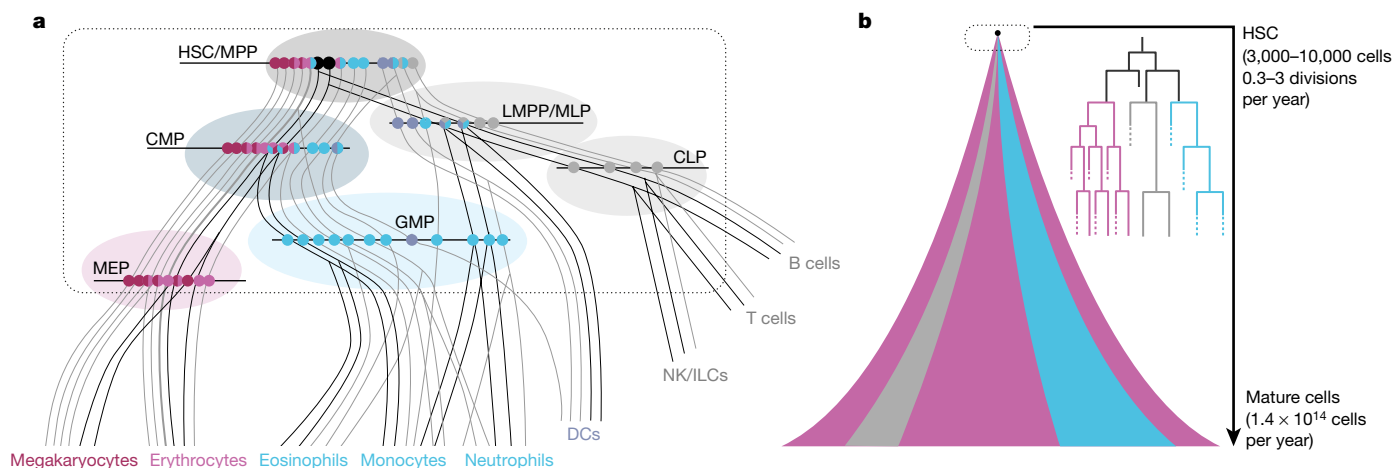


Figure 2 | Trajectory-based visualizations of the haematopoietic hierarchy. **a**, Two-dimensional visualization of early haematopoiesis. Continuous lines denote trajectories of differentiation for different types of single cells present in the phenotypic HSC compartment (grey shaded area). Along these trajectories, cells and their progeny pass through progenitor compartments commonly defined by specific combinations of cell surface markers (shaded areas). Horizontal lines represent snapshots of the lineage potential of the cells present in each phenotypic compartment (single-colour circles denote unilineage cells; two-colour circles denote bilineage cells; three-colours circles denote trilineage cells; black circles denote multipotent cells). In most progenitor compartments, the number of unilineage cells outnumbers that of bi- or trilineage cells. The figure illustrates differentiation trajectories reported in the literature so far, but

their proportions may not reflect the *in vivo* situation. **b**, Three-dimensional visualization of the progeny of a single HSC. Pink, blue and grey represent the erythroid, myeloid and lymphoid lineages, respectively. Cell history, division and progenitor expansion should all be considered when modelling the differentiation journey of one HSC and all of its progeny. In an adult human, there are an estimated 3,000–10,000 HSCs, which probably divide from only once every 3 months to once every 3 years⁵⁸. Humans produce an estimated 1.4×10^{14} mature blood cells per year¹⁴³. The amplification from a few thousand HSCs is therefore staggering, and must include a strong contribution from a transient-amplifying compartment. Also, because there are many more terminally differentiated erythroid cells than myeloid cells, and even less lymphoid cells, all with different turnover rates, the flux into each compartment must be highly regulated.

unique integration events, which serve as barcodes that can be tracked by sequencing. Consequently, the comparison of barcodes in mature lineages after pulse-chase labelling analysis allows reconstruction of single-cell behaviours in native, unperturbed haematopoiesis. In contrast to transplantation approaches, analysis of unperturbed haematopoiesis suggested that (i) MPPs contribute predominantly to the myeloid lineage during the steady state, and (ii) cells functioning as HSCs in transplantation do not have a notable role in steady-state haematopoiesis, which instead seems to be driven almost entirely by cells within the MPP compartment.

An alternative genetic fate-mapping system based on Cre-*loxP*-induced recombination of a transgenic barcode cassette recently achieved temporally controlled barcode induction in single cells, and demonstrated that when HSCs are labelled at the fetal liver stage, their descendants in the adult will mostly contribute to multiple lineages⁹⁷. However, megakaryocytic fate was not analysed, and when the analysis was repeated in adult bone marrow, few barcodes were detected in HSCs as well as mature progeny. Given that each fetal liver HSC divides and therefore give rises to several HSCs in the adult, conclusions about the lineage contribution of individual adult HSCs therefore remained preliminary. Another study from the Camargo group⁶¹ addressed this issue more comprehensively by carrying out a 30-week pulse-chase experiment in adult mice with the sleeping beauty barcode system. 133 barcodes were detected in HSCs and at least one of four mature lineages (megakaryocyte, erythroid, granulocyte or B cell). Interestingly, more than half of these 133 HSC barcodes were present only in megakaryocytes, and only a minority of the remaining barcodes were present in more than one mature lineage. Coupled with analysis at shorter pulse-chase intervals and comprehensive single-cell RNA-seq analysis, this study therefore concluded that during homeostatic unperturbed haematopoiesis, megakaryocytes can arise independently from other lineages, and the phenotypic long-term HSC population as defined by transplantation assays actively contributes to megakaryocyte output.

In the HSC compartment, a transposon tag may often be present in just one or two cells, because HSC clones will rarely amplify during unperturbed haematopoiesis, thus making barcode detection less reliable.

It is therefore noteworthy that both the Rodewald and Reizis groups^{98,99} found that the HSC compartment contributed more to multilineage blood production than what was estimated by the transposon approach. Busch *et al.* also estimated the kinetics with which cells transit through their differentiation trajectories. Interestingly, flux into the lymphoid branch is 180-fold less than in the myeloid branch. Flux into the erythroid lineage was not assessed, but is likely to be even higher than in the myeloid lineage (Fig. 2b). Flux analysis also found substantial self-renewal capacity in the short-term HSC/MPP compartment, consistent with a recent report of long-term normal haematopoiesis in mice in which the HSPC compartment was ablated by 90%¹⁰⁰.

Future approaches are likely to use barcodes that are expressed under a strong promoter, and therefore can be detected reliably by scRNA-seq. This would afford true single-cell resolution for the analysis of clonal relationships and single-cell transcriptomes, offering the exciting possibility of defining the native hierarchy agnostic of sorting strategies that were developed using transplantation. Another pertinent question is to what extent laboratory mice kept under sterile, pathogen-free conditions are a suitable model for human haematopoiesis, which is constantly challenged by exposure to infectious agents, and has to function over a much longer lifespan. Long-term follow-up of patients who underwent autologous transplantation has already revealed previously unknown functional aspects of human haematopoiesis, such as the number, stability and dynamics of individual HSCs over many years¹⁰¹. Background somatic mutations represent unique barcodes that can be exploited to reconstruct clonal lineage relationships^{102,103} and may thus represent another approach to investigate the dynamics of single human HSPCs over extended periods of time.

Haematopoiesis is flexible in space and time

Blood production needs to have the flexibility to adapt to drastic changes of demand, with evidence accumulating that HSC properties and differentiation journeys can adapt. As reviewed elsewhere¹⁰⁴, haematopoietic development in the embryo is complex, with a series of transient haematopoietic waves across several organs (Fig. 3). Importantly, fetal and adult

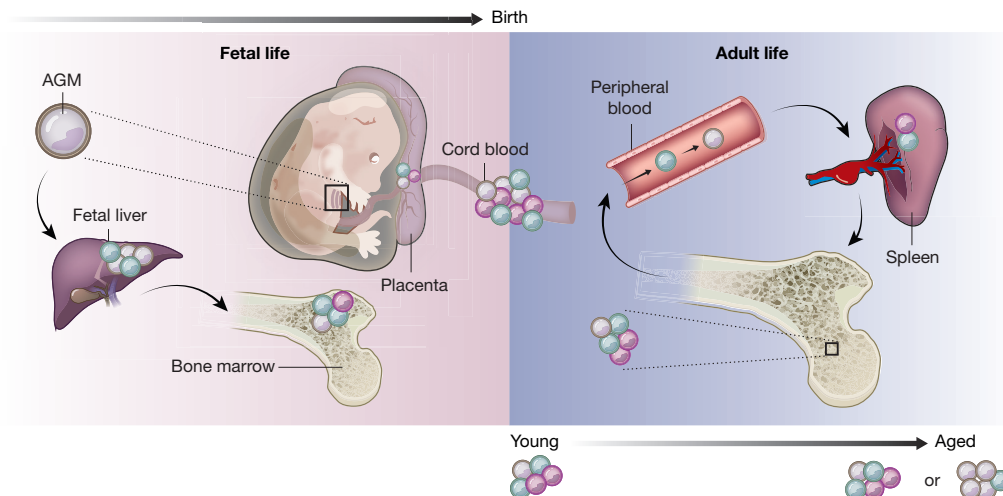


Figure 3 | The composition of the HSPC compartment changes in space and time. HSPCs are found in many organs in the body across a lifetime. Cells of different colours represent distinct HSPC subsets. It is unclear whether all HSPC subsets and differentiation trajectories are present in the same proportions in each of the organs. Current evidence suggests that

age-related changes result from a combination of shifts in the composition of the HSPC pool, as well as phenotypic changes in particular cell types driven by intrinsic genetic or epigenetic changes and systemic alterations of the microenvironment. AGM, aorta gonad mesonephros.

HSCs have fundamentally different regulation and behaviours (reviewed in ref. 105). In mice, there is a switch from a proliferative to a quiescent state between 3 and 4 weeks of age, which coincides with a decrease in self-renewal¹⁰⁶. In humans, the timing of the fetal-to-adult switch differs. When compared to adult bone marrow, human HSCs from neonatal cord blood have increased proliferative potential (as expected from mouse studies), but their cell cycle properties already resemble the adult configuration⁵⁴. Consistently, telomere length in granulocytes, a surrogate for the division rate of HSCs, rapidly declines during the first year of human life¹⁰⁷. There is also evidence that terminally differentiated cells are produced differently during fetal and adult haematopoiesis, because many more single progenitor cells from human fetal liver produce two lineages or more than from adult bone marrow⁷¹. Interestingly, the relative proportion of HSC subsets also change over time with balanced HSC predominating in fetal liver, whereas lymphoid-deficient (also known as myeloid-biased) HSCs accumulate during ageing¹⁰⁸. The effects of ageing on HSCs and more generally on blood production are numerous and have been reviewed elsewhere¹⁰⁹. Changes in both the composition of the HSC pool^{108,110,111} as well as the molecular circuitry of individual HSCs^{34,112–114} can be due to either extrinsic or intrinsic properties, including alterations in the microenvironment, the proliferative history and accumulation of mutations.

The HSC niche is a highly complex ecosystem that sustains HSC function, in particular promoting survival and long-term maintenance of the HSC pool¹¹⁵. If and how niche interactions shape the activity of distinct HSC subsets and the differentiation journeys of HSCs remains unclear. Single-cell transplants and clonal tracking analysis have shown that clones display stereotypical behaviour over serial rounds of transplantation, arguing that their characteristic outputs are not extensively niche-dependent^{15,59,64,65}. However, it is possible that distinct HSC subsets may have different niche preferences. Furthermore, even though the vast majority of HSCs are located in the bone marrow in adults, a small percentage of HSCs are released in the blood with circadian-clock controlled patterns¹¹⁶, and HSCs can also be found in the spleen¹¹⁷ and lungs¹¹⁸ (Fig. 3). The role of these extramedullary niches and whether they host specific subsets of HSCs or influence their differentiation or clonal expansion capacity will have to be explored at single-cell resolution. Finally, many types of stress directly affect HSC function: DNA damage^{40,41}, inflammation¹¹⁹, acute or chronic infection^{120,121}, psychosocial stress¹²², metabolic stress³³ and obesity¹²³. For most of these processes, insights have been gained on the molecular mechanisms that drive the changes in HSC or progenitor cell function. There are also examples that show that

at the cellular level, not all HSC or progenitor subsets respond equally to these stresses. For example, in emergency myelopoiesis, MPP2 and MPP3 drive enhanced production of GMPs^{13,124}, which reorganize themselves spatially and activate a self-renewal network¹²⁴. We are thus only beginning to understand how stress responses can reshape the relative abundance and possibly differentiation trajectories of different HSPC subsets, and less still is known about the cellular and molecular control/feedback mechanisms that maintain and/or re-establish homeostasis.

Implications for human disease

Our understanding of haematopoiesis is currently undergoing several shifts. First, the demarcations between stem and progenitor cells that were previously considered rather rigid are becoming increasingly blurred. Second, cell fate choices upstream of the classically defined bi/oligopotent progenitor cells may be more prevalent than previously thought. Third, the loss of key stem-cell characteristics may be largely decoupled from the initiation of specific lineage differentiation programmes. Finally, measuring cellular fates *in vivo* without the need for highly disruptive transplantation procedures has highlighted a previously underappreciated importance of short-term HSC/MPPs in unperturbed haematopoiesis. These and other revisions of our understanding of haematopoiesis as a stem-cell developmental system have major implications for the diagnosis, prognosis and treatment of haematological diseases.

Haematology as a clinical discipline has a long track record of being an early adopter of the latest technological developments, which recently included the first successful gene therapy trials^{125,126} as well as some of the first comprehensive cancer genome studies¹²⁷. With respect to haematopoietic malignancies, much current research focuses on identifying the cell of origin, which acquired the first somatic mutation within the multistep progression towards a full-blown malignancy. It is widely accepted now that stem and progenitor cells have a major role in the development of myeloid malignancies (chronic myelogenous leukaemia and acute myelogenous leukaemia) and myeloproliferative neoplasms. More recent evidence also implicates HSCs and lymphoid progenitors in the early stages of hairy cell leukaemia¹²⁸ and lymphoma¹²⁹. It is beyond the scope of this Review to list the effects of each of the driver mutations on HSC and progenitor cell function, but it is worth noting that each one of them will reshape the balance of differentiation trajectories and generate complex clonal dynamic patterns. Because cellular context influences the potential effect of leukaemogenic mutations, the newly recognized fluidity of cellular states within the HSPC compartment suggests greater disease heterogeneity between patients, because even when two patients

carry identical founder mutations, the likelihood that they arose in identical cellular states is small. Moreover, the malignant transformation is likely to open up new molecular states and trajectories. In acute myelogenous leukaemia, for example, leukaemic stem cells are defined by a chimaeric transcriptional state^{10,130}, and single-cell proteomic approaches have demonstrated the existence of distinct differentiation trajectories for malignant cells¹³¹.

Given the increasing recognition of discrepancies between the clonal behaviour of native unperturbed HSPCs versus transplantation, xenotransplantation of human leukaemic cells into immunocompromised mice will share similar limitations, in which exposing leukaemic cells to a transplantation assay may induce cellular behaviour that would never occur in a human patient. Exciting prospects may be offered here by new models that may permit transplantation without irradiation¹³² and the use of ossicles templated with human bone marrow stromal cells^{133,134}. It nevertheless seems imperative to invigorate research efforts that directly track disease in human patients. This will require careful design of patient cohorts and new screening technologies, including multiomic single-cell technologies to define the cellular state as well as the mutation burden of individual cells, as demonstrated recently for the gene fusion *BCR-ABL*, found in most patients with chronic myelogenous leukaemia¹³⁵.

Bone marrow transplantation is likely to remain an important curative therapy for many patients with leukaemia, and with progress in the cell and gene therapy field, may find much wider applicability. Because suitable donor material for bone marrow transplantation remains rate limiting, a promising approach is to produce HSCs from other cell types by borrowing regulatory processes known to be important during developmental haematopoiesis, as illustrated by the recent generation of stem cells with long-term engrafting capability from human pluripotent and mouse adult endothelial cells^{136,137}. A better understanding of the HSC state and cell fate decision making will provide new possibilities to develop protocols for *in vitro* amplification of HSCs^{138–140}, and also facilitate optimization of protocols for robust genome engineering of HSCs with vast potential for treating common diseases ranging from inherited red blood cell to autoimmune disorders. Expansion of clinical applications will greatly benefit from a better understanding of the self-renewal and differentiation potential of individual HSPCs, coupled with robust prediction algorithms from molecular profiling data to evaluate the efficacy of cell therapy products. Haematopoiesis is therefore well positioned to lead the way in the gene and cell therapy arena, and we may not be too far away from a future in which haematopoiesis will be firmly established not just as a stem cell, but also as a therapeutic model.

Received 30 July; accepted 8 November 2017.

1. Haeckel, E. *Natürliche Schöpfungsgeschichte* (Georg Reimer, 1868).
2. Pappenheim, A. Ueber Entwicklung und Ausbildung der erythroblasten. *Virchows Arch Pathol Anat* **145**, 587–643 (1896).
3. Pappenheim, A. Zwei Fälle akuter grosslymphozytärer Leukämie. *Fol Haematol* **4**, 301–308 (1907).
4. Jacobson, L. O., Simmons, E. L., Marks, E. K. & Eldredge, J. H. Recovery from radiation injury. *Science* **113**, 510–511 (1951).
5. Till, J. E. & McCulloch, E. A. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat. Res.* **14**, 213–222 (1961).
6. Becker, A. J., McCulloch, E. A. & Till, J. E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452–454 (1963).
7. Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58–62 (1988).
8. Doulatov, S. *et al.* Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat. Immunol.* **11**, 585–593 (2010).
9. Adolfsson, J. *et al.* Identification of Flt3⁺ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* **121**, 295–306 (2005).
10. Goardon, N. *et al.* Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell* **19**, 138–152 (2011).
11. Sanjuan-Pla, A. *et al.* Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232–236 (2013).
This study was one of the first to provide evidence of an HSC differentiation bias along the megakaryocyte-platelet lineage.

12. Yamamoto, R. *et al.* Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell* **154**, 1112–1126 (2013).
This is a comprehensive single-cell transplantation study that demonstrates that single cells can repopulate long-term after transplantation without contributing to all blood lineages.
13. Pietras, E. M. *et al.* Functionally distinct subsets of lineage-biased multipotent progenitors control blood production in normal and regenerative conditions. *Cell Stem Cell* **17**, 35–46 (2015).
This study provides evidence of lineage biases in MPP populations.
14. Cabezas-Wallscheid, N. *et al.* Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell* **15**, 507–522 (2014).
15. Dykstra, B. *et al.* Long-term propagation of distinct hematopoietic differentiation programs *in vivo*. *Cell Stem Cell* **1**, 218–229 (2007).
16. Benveniste, P. *et al.* Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential. *Cell Stem Cell* **6**, 48–58 (2010).
17. Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
18. Müller-Sieburg, C. E., Cho, R. H., Thoman, M., Adkins, B. & Sieburg, H. B. Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. *Blood* **100**, 1302–1309 (2002).
19. Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129 (2008).
20. Foudi, A. *et al.* Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. *Nat. Biotechnol.* **27**, 84–90 (2009).
21. Cross, M. A. & Enver, T. The lineage commitment of haematopoietic progenitor cells. *Curr. Opin. Genet. Dev.* **7**, 609–613 (1997).
22. Chambers, S. M. *et al.* Hematopoietic fingerprints: an expression database of stem cells and their progeny. *Cell Stem Cell* **1**, 578–591 (2007).
23. Novershtern, N. *et al.* Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309 (2011).
24. Laurenti, E. *et al.* The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat. Immunol.* **14**, 756–763 (2013).
25. Chen, L. *et al.* Transcriptional diversity during lineage commitment of human blood progenitors. *Science* **345**, 1251033 (2014).
26. Bock, C. *et al.* DNA methylation dynamics during *in vivo* differentiation of blood and skin stem cells. *Mol. Cell* **47**, 633–647 (2012).
27. Farlik, M. *et al.* DNA methylation dynamics of human hematopoietic stem cell differentiation. *Cell Stem Cell* **19**, 808–822 (2016).
28. Bocker, M. T. *et al.* Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging. *Blood* **117**, e182–e189 (2011).
29. Ji, H. *et al.* Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* **467**, 338–342 (2010).
30. Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
31. Cheshier, S. H., Morrison, S. J., Liao, X. & Weissman, I. L. *In vivo* proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc. Natl Acad. Sci. USA* **96**, 3120–3125 (1999).
32. Cabezas-Wallscheid, N. *et al.* Vitamin A-retinoic acid signaling regulates hematopoietic stem cell dormancy. *Cell* **169**, 807–823 (2017).
This is a comprehensive study that describes the molecular circuitry that maintains dormant HSCs.
33. Warr, M. R. *et al.* FOXO3A directs a protective autophagy program in haematopoietic stem cells. *Nature* **494**, 323–327 (2013).
34. Ho, T. T. *et al.* Autophagy maintains the metabolism and function of young and old stem cells. *Nature* **543**, 205–210 (2017).
35. Simsek, T. *et al.* The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche. *Cell Stem Cell* **7**, 380–390 (2010).
36. Takubo, K. *et al.* Regulation of glycolysis by Pdk functions as a metabolic checkpoint for cell cycle quiescence in hematopoietic stem cells. *Cell Stem Cell* **12**, 49–61 (2013).
37. Vannini, N. *et al.* Specification of haematopoietic stem cell fate via modulation of mitochondrial activity. *Nat. Commun.* **7**, 13125 (2016).
38. Ito, K. *et al.* Self-renewal of a purified Tie2⁺ hematopoietic stem cell population relies on mitochondrial clearance. *Science* **354**, 1156–1160 (2016).
39. Signer, R. A. J., Magee, J. A., Salic, A. & Morrison, S. J. Haematopoietic stem cells require a highly regulated protein synthesis rate. *Nature* **509**, 49–54 (2014).
40. Mohrin, M. *et al.* Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. *Cell Stem Cell* **7**, 174–185 (2010).
41. Milyavsky, M. *et al.* A distinctive DNA damage response in human hematopoietic stem cells reveals an apoptosis-independent role for p53 in self-renewal. *Cell Stem Cell* **7**, 186–197 (2010).
42. van Galen, P. *et al.* The unfolded protein response governs integrity of the haematopoietic stem-cell pool during stress. *Nature* **510**, 268–272 (2014).
43. Mohrin, M. *et al.* Stem cell aging. A mitochondrial UPR-mediated metabolic checkpoint regulates hematopoietic stem cell aging. *Science* **347**, 1374–1377 (2015).
44. van Galen, P. *et al.* Reduced lymphoid lineage priming promotes human hematopoietic stem cell expansion. *Cell Stem Cell* **14**, 94–106 (2014).

73. Miyawaki, K. *et al.* Identification of unipotent megakaryocyte progenitors in human hematopoiesis. *Blood* **129**, 3332–3343 (2017).
74. Sanada, C. *et al.* Adult human megakaryocyte-erythroid progenitors are in the CD34⁺CD38^{mid} fraction. *Blood* **128**, 923–933 (2016).
75. Pronk, C. J. H. *et al.* Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell* **1**, 428–442 (2007).
76. Lee, J. *et al.* Lineage specification of human dendritic cells is marked by IRF8 expression in hematopoietic stem cells and multipotent progenitors. *Nat. Immunol.* **18**, 877–888 (2017).
77. Brady, G., Barbara, M. & Iscove, N. Representative *in vitro* cDNA amplification from individual hematopoietic cells and colonies. *Methods Mol. Cell. Biol.* **2**, 17–25 (1990).
78. Warren, L., Bryder, D., Weissman, I. L. & Quake, S. R. Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proc. Natl. Acad. Sci. USA* **103**, 17807–17812 (2006).
79. Moignard, V. *et al.* Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat. Cell Biol.* **15**, 363–372 (2013).
80. Guo, G. *et al.* Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell* **13**, 492–505 (2013).
81. Pina, C. *et al.* Inferring rules of lineage commitment in haematopoiesis. *Nat. Cell Biol.* **14**, 287–294 (2012).
82. Wilson, N. K. *et al.* Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. *Cell Stem Cell* **16**, 712–724 (2015).
83. Paul, F. *et al.* Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* **163**, 1663–1677 (2015).
- This is a comprehensive single-cell RNA-seq study of the myelo-erythroid progenitor compartment.**
84. Nestorowa, S. *et al.* A single cell resolution map of mouse haematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20–e31 (2016).
- This study provides the first description of the whole haematopoietic hierarchy by single-cell RNA-seq.**
85. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
86. Haghverdi, L., Büttner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015).
87. Grün, D. *et al.* De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* **19**, 266–277 (2016).
88. Setty, M. *et al.* Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* **34**, 637–645 (2016).
89. Olsson, A. *et al.* Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* **537**, 698–702 (2016).
90. Velten, L. *et al.* Human hematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).
- This study couples single-cell RNA-seq and index sorting to delineate differentiation journeys of human HSCs.**
91. Naik, S. H. *et al.* Diverse and heritable lineage imprinting of early hematopoietic progenitors. *Nature* **496**, 229–232 (2013).
92. Lu, R., Neff, N. F., Quake, S. R. & Weissman, I. L. Tracking single hematopoietic stem cells *in vivo* using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat. Biotechnol.* **29**, 928–933 (2011).
93. Cheung, A. M. S. *et al.* Analysis of the clonal growth and differentiation dynamics of primitive barcoded human cord blood cells in NSG mice. *Blood* **122**, 3129–3137 (2013).
94. Dick, J. E., Magli, M. C., Huszar, D., Phillips, R. A. & Bernstein, A. Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of *W/W^u* mice. *Cell* **42**, 71–79 (1985).
95. Lemischka, I. R., Raulet, D. H. & Mulligan, R. C. Developmental potential and dynamic behavior of hematopoietic stem cells. *Cell* **45**, 917–927 (1986).
96. Sun, J. *et al.* Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327 (2014).
97. Pei, W. *et al.* Polylox barcoding reveals haematopoietic stem cell fates realized *in vivo*. *Nature* **548**, 456–460 (2017).
98. Sawai, C. M. *et al.* Hematopoietic stem cells are the major source of multilineage hematopoiesis in adult animals. *Immunity* **45**, 597–609 (2016).
99. Busch, K. *et al.* Fundamental properties of unperturbed haematopoiesis from stem cells *in vivo*. *Nature* **518**, 542–546 (2015).
- In this study, clonal-tracking analysis is coupled with mathematical modelling to define the flux of stem cells into the different lineage branches under unperturbed and transplantation conditions.**
100. Schoedel, K. B. *et al.* The bulk of the hematopoietic stem cell population is dispensable for murine steady-state and stress hematopoiesis. *Blood* **128**, 2285–2296 (2016).
101. Biasco, L. *et al.* In Vivo tracking of human hematopoiesis reveals patterns of clonal dynamics during early and steady-state reconstitution phases. *Cell Stem Cell* **19**, 107–119 (2016).
102. Behjati, S. *et al.* Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**, 422–425 (2014).
103. Biezuner, T. *et al.* A generic, cost-effective, and scalable cell lineage analysis platform. *Genome Res.* **26**, 1588–1599 (2016).
104. Ivanovs, A. *et al.* Human hematopoietic stem cell development: from the embryo to the dish. *Development* **144**, 2323–2337 (2017).

105. Copley, M. R. & Eaves, C. J. Developmental changes in hematopoietic stem cell properties. *Exp. Mol. Med.* **45**, e55 (2013).
106. Bowie, M. B. *et al.* Identification of a new intrinsically timed developmental checkpoint that reprograms key hematopoietic stem cell properties. *Proc. Natl Acad. Sci. USA* **104**, 5878–5882 (2007).
107. Rufer, N. *et al.* Telomere fluorescence measurements in granulocytes and T lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory T cells in early childhood. *J. Exp. Med.* **190**, 157–167 (1999).
108. Benz, C. *et al.* Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. *Cell Stem Cell* **10**, 273–283 (2012).
- This study provides evidence that differentiation biases of HSCs change over a lifetime.**
109. Geiger, H., de Haan, G. & Florian, M. C. The ageing haematopoietic stem cell compartment. *Nat. Rev. Immunol.* **13**, 376–389 (2013).
110. Young, K. *et al.* Progressive alterations in multipotent hematopoietic progenitors underlie lymphoid cell loss in aging. *J. Exp. Med.* <http://doi.org/10.1084/jem.20160168> (2016).
111. Grover, A. *et al.* Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nat. Commun.* **7**, 11075 (2016).
112. Beerman, I. & Rossi, D. J. Epigenetic control of stem cell potential during homeostasis, aging, and disease. *Cell Stem Cell* **16**, 613–625 (2015).
113. Flach, J. *et al.* Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. *Nature* **512**, 198–202 (2014).
114. Sun, D. *et al.* Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell* **14**, 673–688 (2014).
115. Baryawno, N., Severe, N. & Scadden, D. T. Hematopoiesis: reconciling historic controversies about the niche. *Cell Stem Cell* **20**, 590–592 (2017).
116. Méndez-Ferrer, S., Lucas, D., Battista, M. & Frenette, P. S. Haematopoietic stem cell release is regulated by circadian oscillations. *Nature* **452**, 442–447 (2008).
117. Inra, C. N. *et al.* A perisinusoidal niche for extramedullary haematopoiesis in the spleen. *Nature* **527**, 466–471 (2015).
118. Lefrançois, E. *et al.* The lung is a site of platelet biogenesis and a reservoir for haematopoietic progenitors. *Nature* **544**, 105–109 (2017).
119. Takizawa, H., Boettcher, S. & Manz, M. G. Demand-adapted regulation of early hematopoiesis in infection and inflammation. *Blood* **119**, 2991–3002 (2012).
120. Matatall, K. A. *et al.* Chronic infection depletes hematopoietic stem cells through stress-induced terminal differentiation. *Cell Reports* **17**, 2584–2595 (2016).
121. Hirche, C. *et al.* Systemic virus infections differentially modulate cell cycle state and functionality of long-term hematopoietic stem cells *in vivo*. *Cell Reports* **19**, 2345–2356 (2017).
122. Heidt, T. *et al.* Chronic variable stress activates hematopoietic stem cells. *Nat. Med.* **20**, 754–758 (2014).
123. Ambrosi, T. H. *et al.* Adipocyte accumulation in the bone marrow during obesity and aging impairs stem cell-based hematopoietic and bone regeneration. *Cell Stem Cell* **20**, 771–784 (2017).
124. Hérald, A. *et al.* Myeloid progenitor cluster formation drives emergency and leukaemic myelopoiesis. *Nature* **544**, 53–58 (2017).
125. Naldini, L. Gene therapy returns to centre stage. *Nature* **526**, 351–360 (2015).
126. Cavazzana, M., Ribeil, J.-A., Lagresle-Peyrou, C. & André-Schmutz, I. Gene therapy with hematopoietic stem cells: the diseased bone marrow's point of view. *Stem Cells Dev.* **26**, 72–76 (2017).
127. Steensma, D. P. The beginning of the end of the beginning in cancer genomics. *N. Engl. J. Med.* **368**, 2138–2140 (2013).
128. Chung, S. S. *et al.* Hematopoietic stem cell origin of *BRAF*V600E mutations in hairy cell leukemia. *Sci. Transl. Med.* **6**, 238ra71 (2014).
129. Horton, S. J. *et al.* Early loss of *Crebbp* confers malignant stem cell properties on lymphoid progenitors. *Nat. Cell Biol.* **19**, 1093–1104 (2017).
130. Eppert, K. *et al.* Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat. Med.* **17**, 1086–1093 (2011).
131. Levine, J. H. *et al.* Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
132. Cosgun, K. N. *et al.* Kit regulates HSC engraftment across the human-mouse species barrier. *Cell Stem Cell* **15**, 227–238 (2014).
133. Reinisch, A. *et al.* A humanized bone marrow ossicle xenotransplantation model enables improved engraftment of healthy and leukemic human hematopoietic cells. *Nat. Med.* **22**, 812–821 (2016).
134. Sontakke, P. *et al.* Modeling BCR-ABL and MLL-AF9 leukemia in a human bone marrow-like scaffold-based xenograft model. *Leukemia* **30**, 2064–2073 (2016).
135. Giustacchini, A. *et al.* Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat. Med.* **23**, 692–702 (2017).
136. Sugimura, R. *et al.* Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature* **545**, 432–438 (2017).
137. Lis, R. *et al.* Conversion of adult endothelium to immunocompetent haematopoietic stem cells. *Nature* **545**, 439–445 (2017).
138. Wagner, J. E., Jr *et al.* Phase I/II trial of stemregen-1 expanded umbilical cord blood hematopoietic stem cells supports testing as a stand-alone graft. *Cell Stem Cell* **18**, 144–155 (2016).
139. Boitano, A. E. *et al.* Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. *Science* **329**, 1345–1348 (2010).
140. Fares, I. *et al.* Cord blood expansion. *Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal*. *Science* **345**, 1509–1512 (2014).
141. Goodell, M. A., Brose, K., Paradis, G., Conner, A. S. & Mulligan, R. C. Isolation and functional properties of murine hematopoietic stem cells that are replicating *in vivo*. *J. Exp. Med.* **183**, 1797–1806 (1996).
142. Görgens, A. *et al.* Revision of the human hematopoietic tree: granulocyte subtypes derive from distinct hematopoietic lineages. *Cell Reports* **3**, 1539–1552 (2013).
143. Dancay, J. T., Deubelbeiss, K. A., Harker, L. A. & Finch, C. A. Neutrophil kinetics in man. *J. Clin. Invest.* **58**, 705–715 (1976).
144. Hoogenkamp, M. *et al.* Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. *Blood* **114**, 299–309 (2009).
145. Hoppe, P. S. *et al.* Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature* **535**, 299–302 (2016).
146. Rieger, M. A., Hoppe, P. S., Smejkal, B. M., Eitelhuber, A. C. & Schroeder, T. Hematopoietic cytokines can instruct lineage choice. *Science* **325**, 217–218 (2009).
147. Nishikawa, K. *et al.* Self-association of Gata1 enhances transcriptional activity *in vivo* in zebra fish embryos. *Mol. Cell. Biol.* **23**, 8295–8305 (2003).
148. Okuno, Y. *et al.* Potential autoregulation of transcription factor PU.1 by an upstream regulatory element. *Mol. Cell. Biol.* **25**, 2832–2845 (2005).
149. Pimanda, J. E. *et al.* Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc. Natl Acad. Sci. USA* **104**, 17692–17697 (2007).
150. Narula, J., Smith, A. M., Gottgens, B. & Ighoshin, O. A. Modeling reveals bistability and low-pass filtering in the network module determining blood stem cell fate. *PLOS Comput. Biol.* **6**, e1000771 (2010).
151. Swiers, G., Patient, R. & Loose, M. Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. *Dev. Biol.* **294**, 525–540 (2006).

Acknowledgements We thank D. Kent for critical reading of the manuscript. E.L. is supported by a Sir Henry Dale fellowship from the Wellcome Trust (WT)/Royal Society. Research in the Laurenti and Gottgens laboratories is supported by the WT, CRUK, Bloodwise, MRC, BBSRC, NIH-NIDDK, and core support grants by the WT and MRC to the Wellcome-MRC Cambridge Stem Cell Institute.

Author Contributions E.L. and B.G. contributed equally to the writing and editing of the manuscript as well as to figure preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to E.L. (el422@cam.ac.uk) or B.G. (bg200@cam.ac.uk).

Skin microbiota–host interactions

Y. Erin Chen^{1,2}, Michael A. Fischbach² & Yasmine Belkaid^{3,4}

The skin is a complex and dynamic ecosystem that is inhabited by bacteria, archaea, fungi and viruses. These microbes—collectively referred to as the skin microbiota—are fundamental to skin physiology and immunity. Interactions between skin microbes and the host can fall anywhere along the continuum between mutualism and pathogenicity. In this Review, we highlight how host–microbe interactions depend heavily on context, including the state of immune activation, host genetic predisposition, barrier status, microbe localization, and microbe–microbe interactions. We focus on how context shapes the complex dialogue between skin microbes and the host, and the consequences of this dialogue for health and disease.

The skin's outermost aspect consists of a lipid- and protein-laden cornified layer dotted with hair follicles and glands that secrete lipids, antimicrobial peptides, enzymes, salts, and many other compounds^{1–6} (Fig. 1a). Whereas the skin surface is an acidic, high-salt, dessicated, aerobic environment, the invaginations that form folliculo-sebaceous units are comparatively anaerobic and even more lipid-rich^{7,8} (Fig. 1b). The skin surface and follicles are physically and chemically distinct from another microbe-rich barrier site: the small and large intestines. The intestine is moist, polysaccharide-rich, neutral in pH, and full of diverse carbon and nitrogen sources^{9–11}. Additionally, in contrast to the hair follicle, deeper aspects of intestinal crypts that are closer to the epithelium become more aerobic while the lumen is more anaerobic^{12,13}. The skin, on the other hand, is replete in diverse and unusual lipids not found elsewhere in the body^{14,15} (Fig. 2). Some of these lipids, such as sapienic acid, can have antimicrobial activities¹⁶, while others, such as triglycerides, can be metabolized by microbes¹⁷ into free fatty acids and di- and monoglycerides that can be bioactive against other microbes or stimulatory to host cells¹⁸.

Across skin regions, the density and variety of glands and hair follicles vary considerably, creating a complex physical and chemical landscape of geographically distinct niches for bacterial growth. For example, *Cutibacterium* (formerly *Propionibacterium*)¹⁹ and *Staphylococcus* species dominate sebaceous areas (such as the face and torso), while *Corynebacterium*, *Staphylococcus*, and beta-Proteobacteria are found in moist areas (such as the armpits and the elbow and knee creases)²⁰.

In broad terms, the chemistry of a skin niche drives its microbiome composition, but unknown microbial and host factors contribute to important species- and strain-level differences in composition. For some species, such as *Cutibacterium acnes*¹⁹, the same strain tends to colonize multiple body sites of the same individual; others, such as *Staphylococcus epidermidis*, differ among body sites of an individual (but tend to be similar in, for example, the axillae of different individuals)²¹. Most metagenomic cataloguing of the human microbiome has focused on species composition. However, recent work demonstrates that, even within the same species, different strains can differ markedly in their effects on the host²². Strain-level differences have been largely unexplored and remain a frontier for studies of the skin microbiota.

The process of skin microbiota assembly begins during birth and proceeds primarily according to body site over several weeks²³. The microbiota shifts notably during puberty, with increased predominance of *Corynebacterium* and *Cutibacterium* (formerly *Propionibacterium*) and decreased abundance of Firmicutes (including *Staphylococcus* and *Streptococcus* species)²⁴. In adulthood, despite the skin's continuous

exposure to the environment, the microbial composition remains surprisingly stable over time²⁵. This suggests that stabilizing, mutually beneficial interactions exist among commensal microbes and between microbes and the host.

The composition of the skin microbiome can shift markedly during inflammation²⁶. It is not yet understood how pathogens and skin inflammation contribute to a vicious cycle, how homeostasis is re-established, or how pathogens interact with the existing commensal population. The critical role of context to the outcome of a microbe–host interaction animates this review. For example, pathogens such as *Staphylococcus aureus* often colonize the skin asymptotically, whereas mutualists such as *S. epidermidis* can at times promote disease^{27,28}. In this Review, we highlight recent work demonstrating that host–microbe interactions fall along a continuum in which pure pathogenicity and mutualism are at extreme ends, and are rarely useful descriptors. We discuss the importance of context—genetic predisposition, the level of host immune activation, the physical and chemical landscape of the niche, and mitigating or activating microbe–microbe interactions—to the outcome of a host–microbe interaction, and consider how colonization extends from the skin's surface into the follicles and even into subcutaneous tissues.

Host–mutualist interactions

Most microbes living on the skin behave as commensal or mutualistic under steady-state conditions. In contrast to the gut of germ-free mice, which shows grossly altered lymphoid organ development, the skin of germ-free mice does not show marked morphological defects^{29,30}. Nonetheless, skin-resident microbes play important roles in the maturation and homeostasis of cutaneous immunity. The skin microbiota modulate the expression of various innate factors, including interleukin 1a (IL-1 α)²⁹; components of complement³¹; and antimicrobial peptides (AMPs), which are produced by keratinocytes and sebocytes (Fig. 1a). Skin-derived AMPs constitute a diverse array of protein families, but cathelicidins and β -defensins predominate. Although some AMPs are constitutively expressed, others can be stimulated by specific members of the microbiota such as *Cutibacterium*^{5,32} or produced by microbes themselves (including *Cutibacterium* thiopeptides³³ and *S. epidermidis* AMPs^{34,35}). It is not yet known how the combination of microbiota-induced and microbiota-produced AMPs shape microbial communities, but this multidirectional signalling is likely to play an important role in the ecology of skin microbial communities.

One major genus of skin-resident bacteria is *Corynebacterium*, members of which are present at all body sites and dominate in moist sites. Interestingly, corynebacteria share many microbiological features with

¹Department of Dermatology, University of California San Francisco, San Francisco, California, USA. ²Department of Bioengineering and ChEM-H, Stanford University, Stanford, California, USA.

³NIAID Microbiome Program, National Institute of Allergy and Infectious Disease, NIH, Bethesda, Maryland, USA. ⁴Mucosal Immunology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Disease, NIH, Bethesda, Maryland, USA.

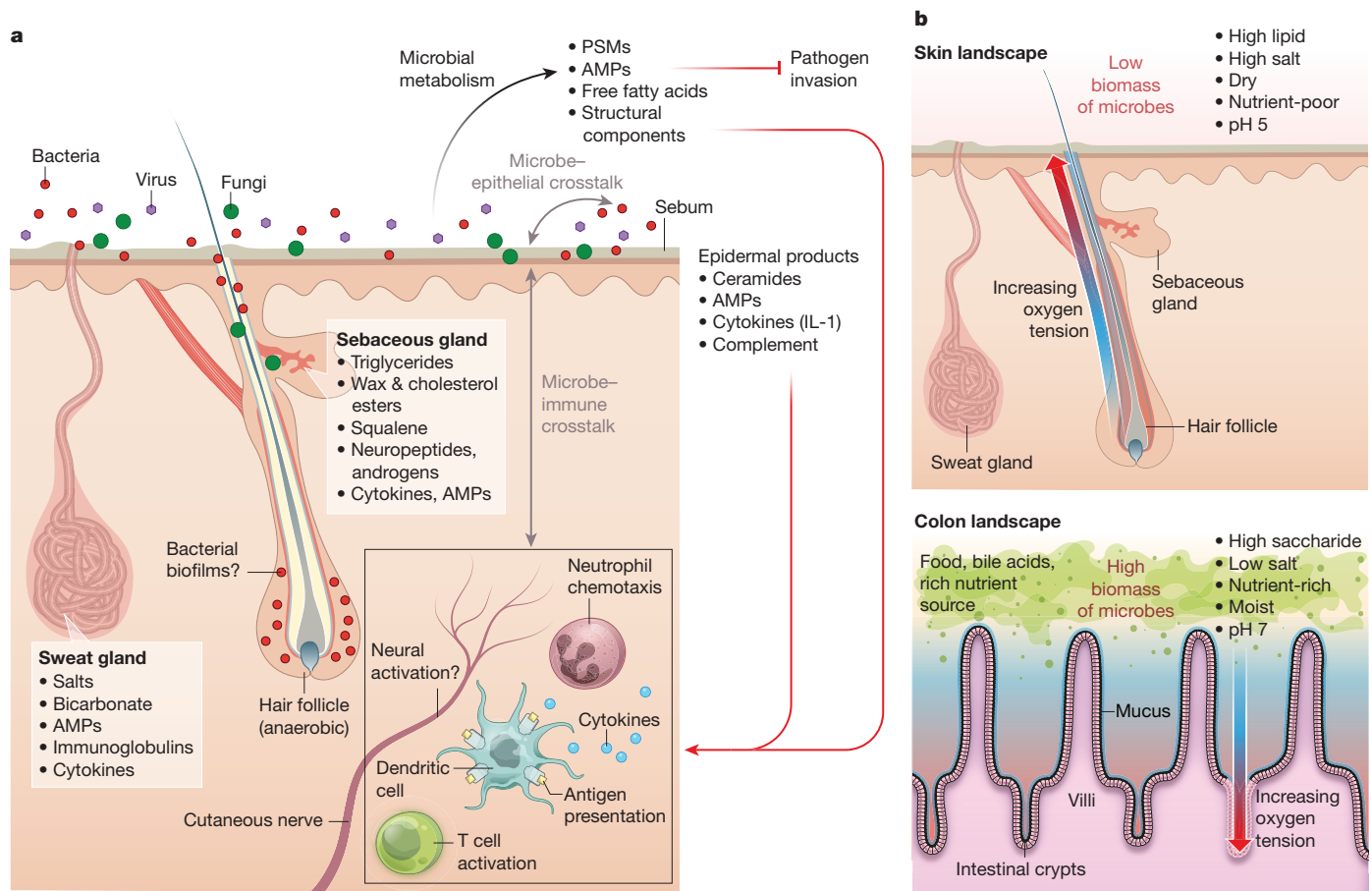


Figure 1 | Crosstalk between skin microbiota and the host. a, Diverse microbes (viruses, fungi and bacteria) cover the skin surface and associated structures (hair follicles, sebaceous glands and sweat glands), possibly forming biofilms at some sites. These microbes metabolize host proteins and lipids and produce bioactive molecules, such as free fatty acids, AMPs, phenol-soluble modulins (PSMs), cell wall components, and antibiotics^{158,159}. These products act on other microbes to inhibit pathogen invasion, on the host epithelium to stimulate keratinocyte-derived immune mediators such as complement and IL-1, and on immune cells in the epidermis and dermis. In turn, host products and immune cell activity influence microbial composition on the skin. **b**, The skin differs from

the closely related mycobacteria, but these two genera interact very differently with the host. It remains a challenge to understand how the immune system distinguishes between bacteria with such similar surface and cellular structures (Fig. 3), and to determine which factors unique to *Corynebacterium* might be responsible for its commensalism. These questions will help to define the molecular-level differences between mutualism and pathogenicity, and explain how commensal bacteria ‘educate’ the cutaneous immune system.

Corynebacterium minutissimum (erythrasma) and *Corynebacterium tenuis* (trichomycosis axillaris) have been associated with superficial skin pathology, but most *Corynebacterium* species present in surveys of the skin microbiome do not cause any known disease. *Corynebacteria* and mycobacteria share the unusual feature among Gram-positive bacteria of having an outer membrane, analogous to that of Gram-negative bacteria (Fig. 3). This outer membrane consists of an outer lipid bilayer of long α -branched fatty acids called mycolic acids, which envelop (and are covalently linked to) the meshwork of peptidoglycan underneath. The *Corynebacterium* cell wall features additional lipoglycans termed lipomannans and lipoarabinomannans, which are anchored to the plasma membrane and have long oligosaccharide chains that emanate from the cell surface. Lipomannans and lipoarabinomannans are ligands for host glycan receptors such as Toll-like receptors (TLRs)

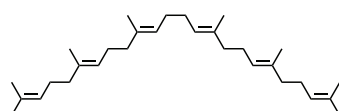
the gut in its physical and chemical properties. The skin is a dry, acidic, lipid-rich, high-salt environment without exogenous nutrient sources, and therefore has low microbial biomass. By contrast, the gut is moist and has abundant nutrients and a thick layer of mucin^{9,159}, enabling it to support much greater microbial biomass. While hair follicles become more anaerobic deeper into the follicle, crypts become more aerobic closer to the epithelium^{7,11}. In addition, material within crypts regularly exchanges with material in the gut lumen, owing to peristalsis, whereas hair follicles have narrow openings filled with sebum and keratinocytic debris, making them more isolated.

and C-type lectin receptors, driving either pro- or anti-inflammatory responses depending on their structure and the immunological context in which they are sensed^{36–42}. In mycobacteria, lipomannans and lipoarabinomannans play important roles in immune recognition and evasion⁴³. It remains to be determined whether similar structures in corynebacteria engage cutaneous immune cells, and whether immune recognition of corynebacteria may protect against future mycobacterial infections.

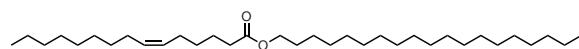
In addition to microbe–host interactions, many reports have suggested that microbe–microbe interactions also impact human health. For example, a common skin resident, *Corynebacterium accolens*, was recently shown to inhibit the growth of *Streptococcus pneumoniae*, a common respiratory tract pathogen⁴⁴. The active principle of this interaction is an essential corynebacterial lipase that hydrolyses triolein to release oleic acid, which inhibits pneumococcal growth⁴⁴. Another common skin resident, *Corynebacterium striatum*, shifts the global transcriptional program of co-cultured *S. aureus* in a way that suppresses virulence-related genes and stimulates genes associated with commensalism²⁸. These data suggest that the role of skin-resident microbes goes beyond competitive exclusion; these microbes are likely to engage in a web of microbe–microbe interactions that help to tune the behaviour of their co-residents in subtle and context-specific ways.

Lipids (15%)

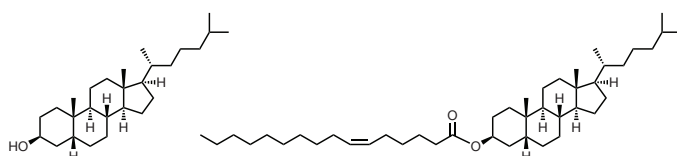
Sebacic lipids: **dry weight (%)**
 Epidermal lipids: **dry weight (%)**



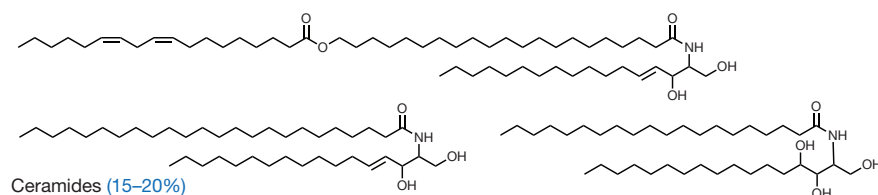
Squalene (12–20%)



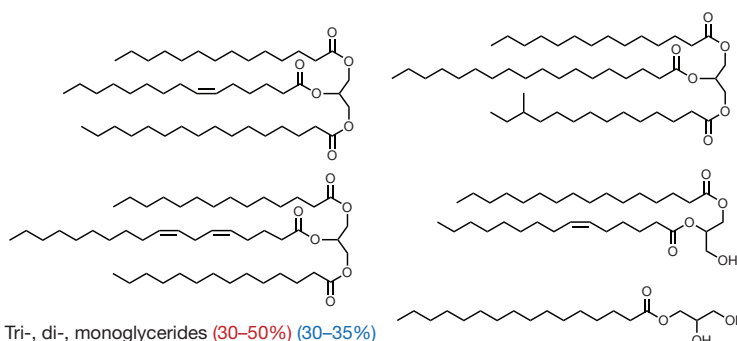
Wax esters (26–30%)



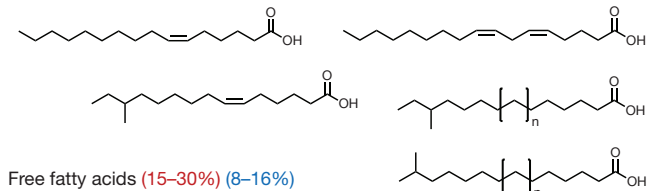
Cholesterol, cholesterol esters (1.5–2.5%) (20–25%)



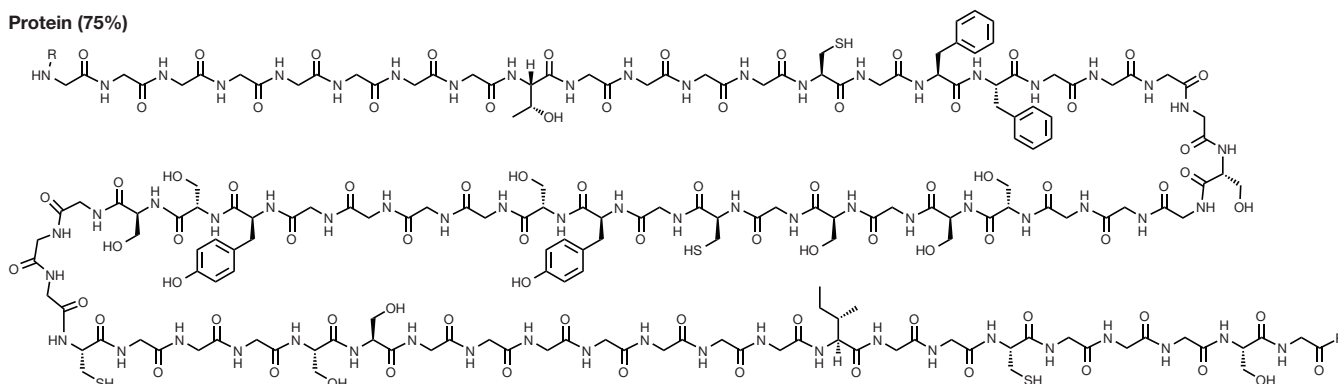
Ceramides (15–20%)



Tri-, di-, monoglycerides (30–50%) (30–35%)



Free fatty acids (15–30%) (8–16%)

Protein (75%)

Representative sequence of loricrin (>70% stratum corneum protein content):

...GGGGGGGGTGGGGCGFFGGGGSGGSGSGCGYSGGGYSGGGCGGGSSGGGGGGIGGCGGGSG...

Figure 2 | Chemistry of the skin. The skin surface consists of a highly organized basketweave structure of keratinocytic proteins and lipids, which are produced by keratinocytes (epidermal lipids) and sebaceous glands (sebaceous lipids)¹⁶⁰. Ceramides are unique to epidermal origin, whereas squalene and wax esters are unique to sebaceous origin, with variable composition under endocrine control¹⁴. Other dominant skin

lipids are cholesterol, triglycerides and free fatty acids (which are often microbial products). Some lipids, such as sphingosine and free fatty acids, demonstrate antimicrobial activity against bacteria, fungi, and viruses⁶⁶ and may have immunomodulatory effects¹⁴⁴. Of keratinocytic proteins, more than 70% of the dry protein weight consists of loricrin, a glycine-rich protein that is thought to have important barrier properties¹⁶⁰.

Another dominant group of skin colonists are the coagulase-negative *Staphylococcus* species, the most prominent of which is *S. epidermidis*. Although *S. epidermidis* can be an opportunistic pathogen in the context of primary or iatrogenic immunosuppression, it functions predominantly as a mutualist. Skin-resident *Staphylococcus* species engage in microbe–microbe interactions that are beneficial to the host. For example, *S. epidermidis* and *Staphylococcus hominis* have been shown to secrete antimicrobial peptides that kill *S. aureus*, and transplantation of these species onto the skin of patients with atopic dermatitis led to decreased colonization by *S. aureus*³⁵.

Recent studies of *S. epidermidis* provided the first evidence that skin-resident bacteria are not just passive residents; they actively engage host immunity through an intact skin barrier, and activate specific immune cell populations in a species- and strain-dependent manner^{29,45}. For instance, some strains of *S. epidermidis* induce activation of

S. epidermidis-specific IL-17⁺CD8⁺ T cells that protect against cutaneous infections by inducing keratinocytes to produce AMPs, a phenomenon called heterologous protection⁴⁵. In addition to their protective role, these commensal-specific T cells also promote wound repair⁴⁶. Of interest, *S. epidermidis* can elicit T cell responses restricted to non-classical major histocompatibility complex (MHC) class I molecules⁴⁶. Thus, non-classical MHC class I molecules, an evolutionarily ancient arm of the immune system, may play an important role in promoting homeostatic immunity to the microbiota. These data show that skin-resident bacteria can have myriad effects on the host; in addition to promoting immune barrier responses, commensal–immune interactions can also affect epithelial biology. The effects of commensal–immune interactions on many other cutaneous processes, including adnexal development, tumorigenesis, ageing, and sensory nerve function, remain to be determined. Additionally, whether immune responses against the skin

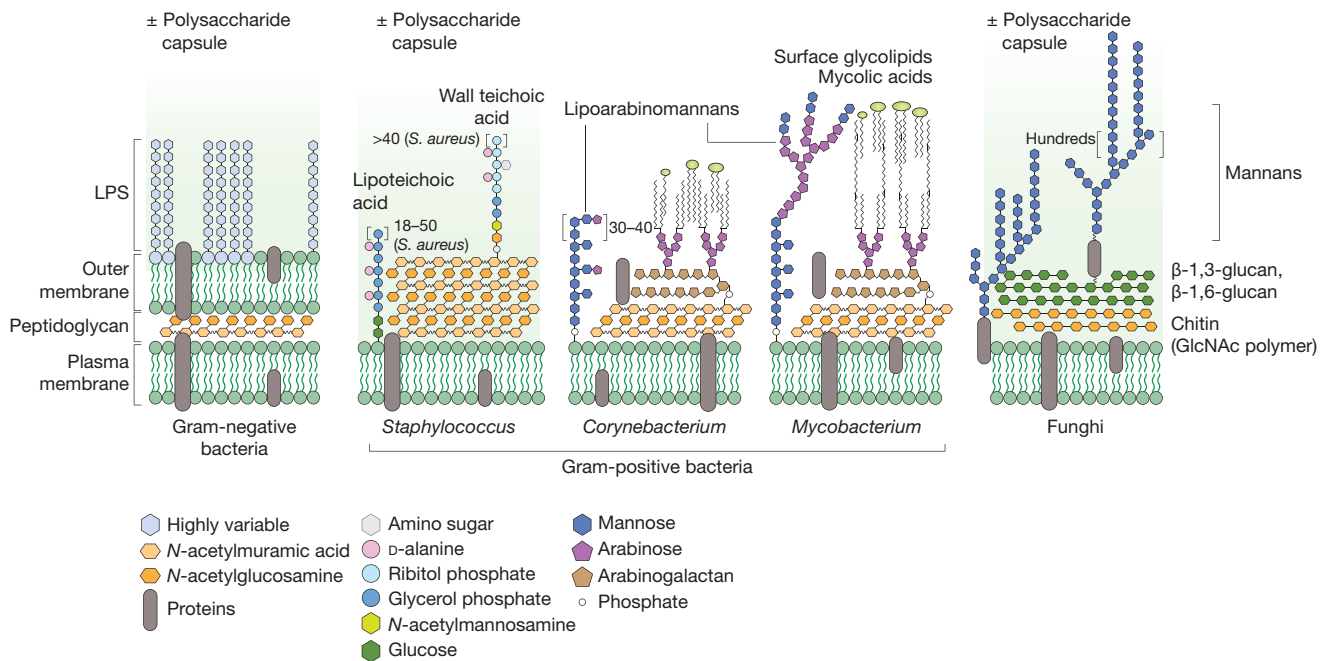


Figure 3 | Chemistry of microbial surfaces. Bacteria and fungi have diverse cell envelopes loaded with immunogenic molecules. Gram-negative bacteria (left) have two lipid bilayers separated by a peptidoglycan cell wall. The outer leaflet of the outer membrane is studded with an immunogenic lipoglycan called lipopolysaccharide (LPS), which has a lipid anchor and highly variable polysaccharide region called the O-antigen. In *Escherichia coli*, for example, 184 different O-antigen structures are known¹⁶¹. Gram-positive bacteria (middle) lack an outer lipid bilayer but have a thicker peptidoglycan cell wall. *Staphylococcus* species have wall-bound and membrane-anchored teichoic acids, which

are analogous to Gram-negative LPS, with a host-facing strain-specific branched polysaccharide. *Corynebacterium* and *Mycobacterium* species also have complex lipoglycans, called lipoteichoic acids. They also have a unique lipid outer layer made up of cell-wall-bound mycolic acids and other noncovalently bound glycolipids. Like Gram-positive bacteria, fungi (right) have only one lipid bilayer membrane. This is covered by a cell wall usually consisting of chitin and a β -D-glucan mesh. The outer layer of the fungal cell wall often contains heavily mannosylated proteins and sometimes a capsule of various polysaccharides. GlcNAc, N-acetylglucosamine.

microbiota also influence microbiota composition or function remains unexplored. In most settings, the skin flora controls skin immunity in an autonomous manner and independently of the gut flora²⁹. This compartmentalization and specialization of responses may have evolved as a mechanism to constrain the adjuvant properties of commensals and unwanted consequences associated with systemic increases in inflammatory responses.

Notably, adaptive responses to members of the skin community develop in the absence of inflammation⁴⁵, in contrast to the response to invading pathogens. This process, termed 'homeostatic immunity', may be induced (at least in part) by the endogenous network of skin-resident antigen-presenting cells⁴⁵. Under steady-state conditions, the skin is populated by highly diverse T cells⁴⁷. Thus, because of the extraordinary number of potential antigens expressed by the microbiota, a substantial fraction of these skin-resident T cells are expected to be microbiota-specific. As a result, primary exposure to a pathogen in the skin or exposure during an injury is likely to occur in the context of a much broader recall response against diverse microbial antigens. The consequences of this phenomenon for tissue responses remain unclear.

Although B cell dynamics in the skin and the role of antibodies in controlling skin microbes are not well understood, it is known that IgA is secreted on the skin surface by eccrine and sebaceous glands^{48,49}. In the gut, IgA has substantial effects on microbiota composition^{50–52} through a process that involves coating commensal bacteria⁵³; in turn, commensal microbes are essential to development of this antibody response^{54,55} and may protect against autoimmunity⁵⁶. Skin commensals are also likely to affect the B cell repertoire, but the extent of this interaction and its effects on the microbiota are not yet known.

In light of the finding that *S. epidermidis* potently and specifically activates a unique branch of adaptive immunity, one major challenge is to dissect mechanistically how specific host cells and receptors recognize

molecular features of *S. epidermidis*. Staphylococci produce a variety of immune modulatory molecules, such as teichoic acids, capsular polysaccharides^{57,58}, and dipeptide aldehydes^{59,60}. Just as *Corynebacterium* and *Mycobacterium* share features but can be distinguished by the host, *S. epidermidis* shares many of these molecular features with the contextual pathogen *S. aureus*. Further studies of how, at the molecular level, *S. aureus* differs from *S. epidermidis* will help us understand how these two important human skin residents are distinguished by the immune system, and will highlight key host targets for more effective and specific therapeutic approaches.

Microbial sensing by the immune system is also likely to be controlled by the developmental stage of the host. Although little is understood about the factors that regulate the acquisition of skin microbes at birth, regulatory T cells, which are highly enriched in the skin tissue, have been proposed to control early dialogue with the microbiota. Indeed, colonization of mouse skin with *S. epidermidis* early in life (but not later) induces tolerance to the same microbe in adulthood⁶¹ and promotes accumulation of *S. epidermidis*-specific regulatory T cells in neonatal skin⁶².

As well as modulating immune cells, *S. epidermidis* and other commensals promote epithelial integrity, especially during tissue repair. For example, an *S. epidermidis* cell wall component mitigates inflammation by binding to TLR2, limiting tissue damage and promoting wound healing⁶³. Other commensal microbiota are also likely to contribute to wound healing, which is a dynamic process associated with global shifts in the skin microbiome; wound bed microbiomes that fail to shift are associated with chronic ulcers^{64,65}. Within chronic wounds, fungi and bacteria form mixed biofilms and certain fungal taxa, such as the phylum *Ascomycota*, are predictive of wounds that take more than eight weeks to heal. The fungal mycobiome and its interactions with commensal bacteria may therefore be important contributors to chronic wounds, via mechanisms that have not been explored^{64,65}.

The skin microbiota are likely to affect many immune-related and immune-independent properties of epithelial health that are not yet appreciated. In the skin, epithelial homeostasis is a constant, active, and energy-intensive process that involves the secretion of complex lipids for signalling and barrier purposes^{2,66}, maintenance of tight junctions^{67,68} and production of a lipid–protein coat to prevent trans-epidermal water loss^{69,70}, repair of UV-mediated and oxidative damage to epithelial cells to prevent malignant transformation^{71,72}, and constant remediation of accidental trauma (for example, scrapes, cuts, and nicks). Any disruption of even a small component of these complex processes can result in extreme phenotypes, such as ichthyoses, blistering disorders, progerias, and diffuse fibrosis^{69,73,74}. How the skin community contributes to these processes remains to be addressed.

Host–pathogen interactions

Microbe–host interactions that drive (or result from) infectious processes have historically received the most investigative attention. A canonical host–pathogen interaction in the skin involves a one-to-one mapping of microbe to disease and an easily identified phenotype of inflammation. Most of what is known about the skin immune system has been discovered by studying interactions of this sort, highlighting the utility of this simplistic paradigm and foreshadowing its limitations. As we will discuss, most microbe–host interactions on the skin are more nuanced; a threshold example is the observation that traditional pathogens often reside on the skin surface in an asymptomatic manner.

In terms of cost and prevalence, one of the most important pathogens of the skin is *S. aureus*. Although more than 30% of healthy individuals are colonized asymptotically by *S. aureus*^{75,76}, it can cause a wide spectrum of infections: some are limited to a single hair follicle (furuncle), others involve subcutaneous tissues (cellulitis), and the most serious feature potentially fatal penetration into any organ in the body, including bone (osteomyelitis), bloodstream (bacterial sepsis), and heart valves (bacterial endocarditis). *S. aureus* has also been implicated in the pathogenesis of chronic diseases such as atopic dermatitis^{22,77–79}, and more recently in systemic lupus erythematosus with renal and skin involvement⁸⁰.

S. aureus is a versatile pathogen with a broad array of virulence factors^{81,82}, including neutrophil-killing toxins⁸³, chemotaxis inhibitors⁸⁴, anti-phagocytic and anti-killing surface molecules^{57,85–88}, superantigens, and immune evasion proteins^{89,90}. In patients with atopic dermatitis, *S. aureus* isolates grow as biofilms on the skin and produce proteases that degrade host AMPs, such as cathelicidin LL-37⁹¹. The host has evolved mechanisms to ward off invasion by *S. aureus* at every level of the skin and subcutaneous tissue. In addition to a diverse arsenal of AMPs covering the epidermis, there is also evidence that adipose tissue under the dermis contributes to the innate immune response. Following breach of the skin barrier and subsequent *S. aureus* infection, local pre-adipocytes proliferate rapidly, expanding the subdermal fat layer and increasing production of the AMP cathelicidin⁹².

Although some virulence and immune evasion elements are conserved across all species of *S. aureus*, there are important strain-level differences. For example, the arginine catabolic mobile element contributes to the ability of USA300, a methicillin-resistant *S. aureus* strain, to thrive in the acidic environment of human skin and resist host polyamines, helping to explain this strain's prevalence in skin and soft tissue infections^{93,94}. Recent work has also shown that certain strains of *S. aureus* are not only associated with more severe atopic dermatitis, but are also sufficient to induce skin inflammation in mice independent of host genetic predisposition²². This work revealed that the most common method of describing microbiome composition, with genus- or species-level data, fails to resolve important functional differences among strains, which can result from modest gene gain or loss events or even differences in gene expression among strains.

Another prominent genus of skin pathogens is *Mycobacterium*, a Gram-positive rod within the phylum *Actinobacteria*. Mycobacteria are a diverse genus of organisms that includes the causative agents of tuberculosis (*Mycobacterium tuberculosis*) and leprosy (*Mycobacterium leprae*), and

other species that cause infections at surgical sites or sites of accidental trauma (for example, *Mycobacterium kansasii*, *Mycobacterium chelonae*, and *Mycobacterium marinum*). *M. tuberculosis*, which generally causes pulmonary or systemic infections, is well known; however, skin and soft tissue infections caused by other *Mycobacterium* species are increasing in prevalence in developed countries⁹⁵ and continue to be serious problems in developing countries⁹⁶. We highlight mycobacteria because they generate an especially broad spectrum of pathologies, from acute systemic illness to skin manifestations to inert granulomas that persist throughout the lifetime of a host. In addition, mycobacteria are closely related to skin-resident corynebacteria but have very different effects on the host. The similarities and differences between *Mycobacterium* and *Corynebacterium* will probably yield insights into a broad swath of fundamentally important host–microbiota interactions.

M. tuberculosis is one of the most successful pathogens on the planet: it colonizes one-quarter of the world's population (1.7 billion people) in the form of a latent infection⁹⁷, with the World Health Organization estimating that 10.4 million new infections occurred in 2015⁹⁸. Among humans who are latent carriers of *M. tuberculosis*, only 10% will suffer reactivation into active tuberculosis during their lifetime⁹⁹. A related pathogen, *M. leprae*, also causes a wide range of diseases, including diverse skin manifestations that can involve the nerves, liver, and bones^{100,101}. The time period between *M. leprae* inoculation and clinical manifestation of infection is typically 2–12 years and can be up to a few decades; during this time, the bacterium handily evades host immunity. It is particularly notable that the majority of humans who harbour *M. tuberculosis* and *M. leprae* neither display obvious pathology nor die from their infection. This suggests that we have much to discover about the mechanisms of long-term immune evasion, and that the traditional definition of 'colonist' may need to expand to include an organism such as *M. tuberculosis*, which on the one hand is a pathogen that kills more than a million people each year¹⁰², and on the other hand lives asymptotically in billions of people and results in the death of only a small percentage of its hosts.

One recently discovered mechanism of mycobacterial host evasion involves the nervous system. *Mycobacterium ulcerans* causes the Buruli ulcer, a progressive, necrotic ulcer that is the third most common mycobacterial disease worldwide¹⁰³. The Buruli ulcer is painless, which contributes to delays in treatment and therefore increases the requirement for more drastic interventions at later infectious stages, when the only available treatment is surgical resection. *M. ulcerans* produces a polyketide toxin, mycolactone, that is essential for virulence. Recent work has shown that mycolactone induces analgesia via the angiotensin II receptor, COX-1, and prostaglandin E₂, ultimately resulting in activation of TRAAK potassium channels and cell hyperpolarization¹⁰⁴. Not only does this work provide possible avenues to therapeutic biomimetics for pain relief and to the development of therapies against *M. ulcerans*, but it also demonstrates a novel mode of pathogen–host interaction in which the peripheral nervous system is targeted directly.

Involvement of the peripheral nervous system may be more general and integral to skin immunity than has been previously recognized. *Candida albicans*, a fungal pathogen, also triggers sensory neurons directly; these neurons then stimulate host immunity and activate protective IL-17-producing dermal T cells¹⁰⁵. *S. aureus* also activates cutaneous neurons directly via N-formylated peptides and the pore-forming toxin α -haemolysin, inducing pain and neuropeptide-mediated induction of vasodilation and inflammation¹⁰⁶. More recently, a direct mechanistic link between neurons and immune cells has been discovered. For instance, in the gut, mucosal neurons were found to produce a neuropeptide, neuromedin U (NMU), that binds an NMU receptor on group 2 innate lymphoid cells (ILC2s) and triggers a protective immune response¹⁰⁷. Direct microbiota–nervous system interactions appear to be a broader phenomenon than was previously appreciated, with examples emerging in other body sites; in a recent case, a microbiota-derived metabolite (isovaleric acid) was shown to trigger a receptor enriched in

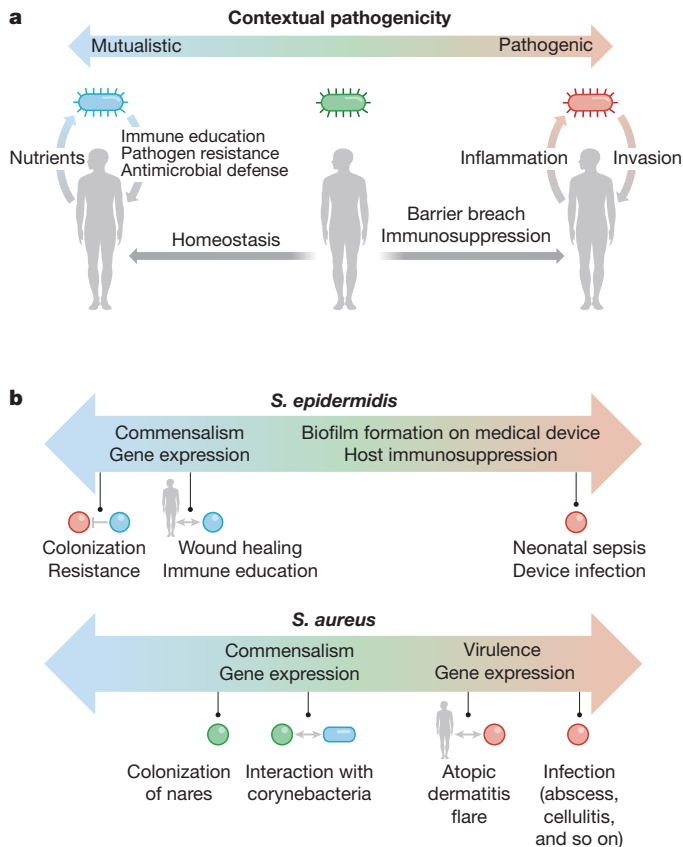


Figure 4 | Contextual pathogenicity. **a**, Microbes exhibit contextual pathogenicity along a spectrum. Host factors such as barrier breaches and immunosuppression bias microbes towards pathogenic behaviour, whereas homeostatic conditions bias them towards mutualistic behaviour. On the microbial side, virulence gene expression and microbe–microbe interactions can also push microbial behaviour to be mutualistic or pathogenic. In a mutualistic host–microbe relationship, the host provides nutrients, while the microbe promotes epithelial and immune homeostasis as well as pathogen resistance through microbial products and occupation of metabolic niches. In a pathogenic relationship, the microbe invades past the epithelium, causing inflammation, and sometimes also benefiting from a host inflammatory response. **b**, Both *S. epidermidis* and *S. aureus* are examples of contextual pathogenicity; *S. epidermidis* is biased towards mutualistic behaviour whereas *S. aureus* displays more pathogenic character.

enterochromaffin cells, resulting in the basolateral release of serotonin, which stimulated sub-epithelial enteric nervous system afferents¹⁰⁸.

As well as evading immunity so that an infection can establish, mycobacteria can also persist for decades inside granulomas, which are organized aggregates of macrophages. Bacteria living within a granuloma, in equilibrium with the host, can be considered a form of tissue colonization. In the gut, for example, certain subsets of commensal bacteria intracellularly colonize CD11c⁺ dendritic cells in healthy mice and promote innate lymphoid cell responses that prevent systemic dissemination of these bacteria, as well as IL-10 production that protects against intestinal inflammation and damage^{109,110}. These data demonstrate that commensal colonists can reside not only on the surface of the host, but also within host cells and tissues, and that these bacteria can elicit a balance of immune-reactive and immunosuppressive responses in the host. For mycobacteria, granulomas may seed the infection of new macrophages, leading to dissemination¹¹¹. However, in 80–90% of healthy patients with latent tuberculosis, reactivation does not occur during the lifetime of the host¹¹². In these patients, *M. tuberculosis* persists within macrophages, using a variety of strategies including efflux pumps that promote drug tolerance^{113–115}, increased cell wall biosynthesis¹¹⁶, and global

transcriptional changes to survive in anaerobic conditions^{117–119}. Within a single host, granulomas can resolve variably: some become sterile, others harbour latent bacteria, and some allow bacteria to escape¹²⁰ in a manner that does not depend on host or bacterial genetic features. Therefore, even for an individual pathogen within a single host, it is not yet understood how the specifics of context affect the outcome of a microbe–host encounter.

As well as bacterial pathogens, there are numerous viral and fungal pathogens of the skin. Some viruses, such as human papillomaviruses and herpesviruses, can cause acute pathology but generally persist in a latent manner for a lifetime. Other viruses, such as the orthopoxvirus vaccinia virus, are cleared after epicutaneous infection. Although our current understanding of these viruses encompasses only their pathogenic behaviour, analyses of the double-stranded DNA virome have shown that papillomaviruses and poxviruses can also colonize hosts asymptomatically^{21,121}. Even in the ‘simple’ case of vaccinia virus, intravital multiphoton microscopy has shown a complex spatial orchestration of immune events; although viral infection occurs in keratinocytes, responding CD8⁺ T cells do not target the infected keratinocytes but rather innate immune cells¹²². Counterintuitively, high local production of the anti-inflammatory cytokine IL-10 helps to limit vaccinia replication and dissemination¹²³.

An immune cell population that has received attention for its ability to promote memory responses to viruses (and, potentially, other members of the microbiota) is resident memory T (T_{RM}) cells—a population of lymphocytes that occupies tissues without recirculating¹²⁴. While most studies have focused on virally induced CD8⁺ T_{RM} cells, CD4⁺ T_{RM} cells (or at least cells with a resident phenotype) have also been shown to accumulate in the skin in response to a large array of infections. Notable differences in anatomical localization exist for memory CD4⁺ and CD8⁺ T cells, with CD4⁺ T cells maintained primarily within the dermis and CD8⁺ T cells within the epidermal compartment¹²⁵. Given that herpesviruses, papillomaviruses and polyomaviruses that infect epithelial cells generate a high burden of chronic skin disease and, in some cases, aggressive skin cancers, there is an urgent need to better understand how these infections are contained by or escape cutaneous immunity.

Contextual pathogenicity

Traditionally, the term ‘pathobiont’ has been applied to organisms that have the potential to cause disease but often colonize a host without inducing pathology. Two prominent gut pathobionts are segmented filamentous bacterium (SFB), which stimulates T helper 17 (T_H17) cells in the mouse gut to confer protection against pathogens but can also induce severe colitis; and *Helicobacter pylori*, which colonizes half of the human population, but in a small percentage can cause peptic ulcer disease and potentiate gastric adenocarcinoma^{126–129}. However, in more extreme contexts of host predisposition, many other microbial species have the potential to cause disease. Patients with primary immunodeficiency (PID), for example, develop chronic, severe skin infections, many of them induced by normal constituents of the microbiota or environmental microbes¹³⁰.

Transitioning from commensalism (for example, on the skin surface or in a follicle) to pathogenicity (for example, in the bloodstream) is a complex and potentially costly process for the microbe. The skin is an imposing physical barrier, and even if it is breached, the microbe must fend off numerous layers of innate responses, including AMPs, proteases, and reactive oxygen species. In addition, would-be pathogens need to induce the expression of genes that enable adhesion, invasion, and immune evasion (that is, virulence factors). As a result, most commensal microbes coexist peacefully with their host, and will exhibit pathogenic potential only in specific settings. Conversely, microbes that are traditionally considered pathogens do not indiscriminately display aggressive behaviour. As discussed above, the term pathogen does not encompass the range of phenotypes displayed by *S. aureus* or *M. tuberculosis*. Viewed in this light, the term pathobiont becomes unhelpfully broad. Therefore, we suggest that the concept of a continuum of contextual pathogenicity and mutualism may be more useful (Fig. 4). In this section,

we discuss examples of microbes that might traditionally be referred to as pathobionts: those that represent the middle of the spectrum between aggressive and mutualistic behaviour.

One example is *C. acnes*, which has been implicated in acne^{131–133} and hidradenitis suppurativa¹³⁴. However, the contribution of *C. acnes* to the pathogenesis of acne is unclear¹³¹. On the one hand, *C. acnes* produces coproporphyrin III, which has been shown to induce the formation of *S. aureus* biofilms¹³⁵—generally seen as a negative consequence for the host. On the other hand, *C. acnes* has also been shown to ferment glycerol into short-chain fatty acids, which suppress the growth of virulent methicillin-resistant *S. aureus* USA300¹³⁶. These data suggest that rich networks of microbe–microbe interactions may govern host inflammation and disease in a strain- and context-dependent manner.

A similar range of harmful and beneficial effects have been demonstrated for other microbes. Herpesviruses are frequent pathogens of the skin, with notable examples including chickenpox (varicella zoster virus) and recurrent labial ulcers (HSV1 and HSV2); however, after the acute infection has resolved, herpesviruses persist within the host as dormant, latent viruses for the host's lifetime. In mice, this herpesvirus latency has been shown to stimulate the immune system in a way that is protective against bacterial pathogens for months^{137,138}. As noted above, *H. pylori* colonizes the stomach and can cause ulcers and promote gastric cancer. However, it has long evolved to live within human hosts, and is also thought to protect against tuberculosis¹³⁹ and allergic diseases such as childhood asthma¹⁴⁰.

Although *S. epidermidis* is generally beneficial to the host, it is also a leading cause of death in premature infants and nosocomial infections^{27,141}. Conversely, *Mycobacterium* species such as *M. leprae* and *M. tuberculosis* are known for their ability to cause serious systemic illness but can persist subclinically inside a granuloma for the lifetime of a host. To generalize, all microbes that reside on or inside a host fall along a spectrum, with some displaying almost no aggressive behaviour and others displaying primarily virulent, invasive phenotypes. The majority of skin residents probably lie somewhere in the middle of the spectrum, and an important challenge in future work will be to better understand which host and environmental factors govern a microbe's switch between passive and aggressive behaviours.

Discussion

The examples discussed herein illustrate that a scheme in which skin-resident microbes are classified as 'full-time' pathogens, mutualists or pathobionts may need to be updated to include the effects of context (Fig. 4). Future work will need to investigate the context-dependent behaviour of resident microbes—how microbe–microbe interactions, host–microbe interactions, and strain-specific differences may govern a microbe's tendency towards cooperation or aggression.

As we learn more about how commensal bacterial strains activate specific immune cell populations, we may be able to harness this specificity by engineering microbes to deliver cytokines, small molecules, or vaccines to specific, activated immune cell populations across an intact skin barrier. A clearer understanding of the dense network of microbe–microbe interactions will also allow us to provide more targeted therapies for dysbiosis, which has been implicated in atopic dermatitis but is also being explored as a pathogenic contributor to many other skin diseases, including psoriasis, hidradenitis suppurativa, and lupus erythematosus. In the gut, using microbes to correct dysbiosis has been successful in the case of faecal transplants for *Clostridium difficile* infections¹⁴², and more recently in the use of a *Lactobacillus plantarum* synbiotic to prevent neonatal sepsis¹⁴³. Similar live microbial therapies for the skin have not yet been developed. However, harnessing the immunomodulatory⁴⁵ or antimicrobial properties of skin commensal bacteria¹⁴⁴ has great potential. Furthermore, because of the unique chemical milieu of the skin (Fig. 2), local alterations in defined nutrients may have a marked impact on the composition or function of skin microbiota and—when rationally designed—could promote the expansion of microbes endowed with regulatory or protective properties.

Another important category of microbe–host interactions that has not been well explored consists of distant effects between microbes at one site, for example in the gut, and host responses at another site, such as the skin. Recently, immune checkpoint blockade has achieved unprecedented success in the treatment of multiple skin cancers that were previously associated with high rates of mortality, including metastatic melanoma, squamous cell carcinoma, and Merkel cell carcinoma^{145–148}. Although these are cutaneous cancers, commensal gut bacteria have been implicated in the efficacy of anti-tumour immunotherapies at distant sites^{149,150}. Conversely, how skin-resident microbes influence immune responses systemically or at distant sites is an important area for further research. Processes that were previously thought to involve skin-limited inflammation, such as plaque psoriasis, have now been linked to an increase in systemic inflammatory co-morbidities, such as atherosclerotic cardiovascular disease^{151–154}. Indeed, many types of immune cell are known to traffic in and out of skin during both homeostasis and inflammation^{155–157}. These observations suggest that the effects of the skin microbiota on the immune system may have wide-ranging systemic sequelae that are ripe for exploration in the near future.

These findings add another layer of complexity to microbe–host interactions, suggesting that research should not only focus on interactions within the local microenvironment but also encompass trafficking of microbiota-educated immune cell populations, or microbial products and metabolites, to other body sites upon stimulation by microbes at diverse barrier sites.

Received 26 September; accepted 28 November 2017.

- De Luca, C. & Valacchi, G. Surface lipids as multifunctional mediators of skin responses to environmental stimuli. *Mediators Inflamm.* **2010**, 321494 (2010).
- van Smeden, J. & Bouwstra, J. A. Stratum corneum lipids: their role for the skin barrier function in healthy subjects and atopic dermatitis patients. *Curr. Probl. Dermatol.* **49**, 8–26 (2016).
- Niyonsaba, F., Kiatsurayanon, C., Chiesilapatham, P. & Ogawa, H. Friends or foes? Host defense (antimicrobial) peptides and proteins in human skin diseases. *Exp. Dermatol.* **26**, 989–998 (2017).
- Bek-Thomsen, M., Lomholt, H. B., Scavenius, C., Enghild, J. J. & Brüggemann, H. Proteome analysis of human sebaceous follicle infundibula extracted from healthy and acne-affected skin. *PLoS One* **9**, e107908 (2014).
- Lee, D.-Y. et al. Sebocytes express functional cathelicidin antimicrobial peptides and can act to kill propionibacterium acnes. *J. Invest. Dermatol.* **128**, 1863–1866 (2008).
- Boncheva, M. The physical chemistry of the stratum corneum lipids. *Int. J. Cosmet. Sci.* **36**, 505–515 (2014).
- Matard, B. et al. First evidence of bacterial biofilms in the anaerobe part of scalp hair follicles: a pilot comparative study in folliculitis decalvans. *J. Eur. Acad. Dermatol. Venereol.* **27**, 853–860 (2013).
- Puhvel, S. M., Reisner, R. M. & Amirian, D. A. Quantification of bacteria in isolated pilosebaceous follicles in normal skin. *J. Invest. Dermatol.* **65**, 525–531 (1975).
- Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2016).
- Peterson, L. W. & Artis, D. Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nat. Rev. Immunol.* **14**, 141–153 (2014).
- Albenberg, L. et al. Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. *Gastroenterology* **147**, 1055–1063.e8 (2014).
- Lind Due, V., Bonde, J., Kann, T. & Perner, A. Extremely low oxygen tension in the rectal lumen of human subjects. *Acta Anaesthesiol. Scand.* **47**, 372 (2003).
- Crompton, D. W. T., Shrimpton, D. H. & Silver, I. A. Measurements of the oxygen tension in the lumen of the small intestine of the domestic duck. *J. Exp. Biol.* **43**, 473–478 (1965).
- Strauss, J. S., Pochi, P. E. & Downing, D. T. The sebaceous glands: twenty-five years of progress. *J. Invest. Dermatol.* **67**, 90–97 (1976).
- Nicolaides, N. Skin lipids: their biochemical uniqueness. *Science* **186**, 19–26 (1974).
- Drake, D. R., Brogden, K. A., Dawson, D. V. & Wertz, P. W. Thematic review series: skin lipids. Antimicrobial lipids at the skin surface. *J. Lipid Res.* **49**, 4–11 (2008).
- Puhvel, S. M., Reisner, R. M. & Sakamoto, M. Analysis of lipid composition of isolated human sebaceous gland homogenates after incubation with cutaneous bacteria. Thin-layer chromatography. *J. Invest. Dermatol.* **64**, 406–411 (1975).

This study showed that common skin bacteria, such as *Cutibacterium* (*Propionibacterium*) species and *Staphylococcus epidermidis*, can modify skin lipids through hydrolysis of triglycerides and esterification of cholesterol, and that these enzymatic activities can be modified by other skin features, such as pH.

18. Sanford, J. A. *et al.* Inhibition of HDAC8 and HDAC9 by microbial short-chain fatty acids breaks immune tolerance of the epidermis to TLR ligands. *Sci. Immunol.* **1**, eaah4609 (2016).
19. Scholz, C. F. P. & Kilian, M. The natural history of cutaneous propionibacteria, and reclassification of selected species within the genus *Propionibacterium* to the proposed novel genera *Acidipropionibacterium* gen. nov., *Cutibacterium* gen. nov. and *Pseudopropionibacterium* gen. nov. *Int. J. Syst. Evol. Microbiol.* **66**, 4422–4432 (2016).
20. Grice, E. A. *et al.* Topographical and temporal diversity of the human skin microbiome. *Science* **324**, 1190–1192 (2009).
Using 16S ribosomal RNA sequencing, this study provided a metagenomic analysis of the human skin microbiome and described previously unappreciated bacterial diversity at different skin sites.
21. Oh, J. *et al.* Biogeography and individuality shape function in the human skin metagenome. *Nature* **514**, 59–64 (2014).
22. Byrd, A. L. *et al.* *Staphylococcus aureus* and *Staphylococcus epidermidis* strain diversity underlying pediatric atopic dermatitis. *Sci. Transl. Med.* **9**, eaal4651 (2017).
This study highlights the utility of shotgun metagenomic sequencing over 16S ribosomal RNA sequencing to assess how strain differences within the same *Staphylococcus epidermidis* species can contribute to disease.
23. Chu, D. M. *et al.* Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat. Med.* **23**, 314–326 (2017).
24. Oh, J., Conlan, S., Polley, E. C., Segre, J. A. & Kong, H. H. Shifts in human skin and nares microbiota of healthy children and adults. *Genome Med.* **4**, 77 (2012).
25. Oh, J., Byrd, A. L., Park, M., Kong, H. H. & Segre, J. A. Temporal stability of the human skin microbiome. *Cell* **165**, 854–866 (2016).
26. Kong, H. H. *et al.* Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res.* **22**, 850–859 (2012).
This study was one of the first to use metagenomic sequencing to characterize dysbiosis in inflammatory skin diseases, showing that atopic dermatitis flares are associated not only with blooms of *Staphylococcus aureus* but also with significant decreases in overall skin microbial diversity.
27. Otto, M. *Staphylococcus epidermidis*—the ‘accidental’ pathogen. *Nat. Rev. Microbiol.* **7**, 555–567 (2009).
28. Ramsey, M. M., Freire, M. O., Gabrilski, R. A., Rumbaugh, K. P. & Lemon, K. P. *Staphylococcus aureus* shifts toward commensalism in response to *Corynebacterium* species. *Front. Microbiol.* **7**, 1230 (2016).
29. Naik, S. *et al.* Compartmentalized control of skin immunity by resident commensals. *Science* **337**, 1115–1119 (2012).
This study demonstrated that skin-resident commensal bacteria are critical for establishing skin immune homeostasis and that this process occurs through an intact, uninfamed skin barrier.
30. Cebra, J. J. Influences of microbiota on intestinal immune system development. *Am. J. Clin. Nutr.* **69**, 1046S–1051S (1999).
31. Chehoud, C. *et al.* Complement modulates the cutaneous microbiome and inflammatory milieu. *Proc. Natl Acad. Sci. USA* **110**, 15061–15066 (2013).
32. Nagy, I. *et al.* *Propionibacterium acnes* and lipopolysaccharide induce the expression of antimicrobial peptides and proinflammatory cytokines/chemokines in human sebocytes. *Microbes Infect.* **8**, 2195–2205 (2006).
33. Christensen, G. J. M. *et al.* Antagonism between *Staphylococcus epidermidis* and *Propionibacterium acnes* and its genomic basis. *BMC Genomics* **17**, 152 (2016).
34. Cogen, A. L. *et al.* Selective antimicrobial action is provided by phenol-soluble modulins derived from *Staphylococcus epidermidis*, a normal resident of the skin. *J. Invest. Dermatol.* **130**, 192–200 (2010).
35. Nakatsuji, T. *et al.* Antimicrobials from human skin commensal bacteria protect against *Staphylococcus aureus* and are deficient in atopic dermatitis. *Sci. Transl. Med.* **9**, eaah4680 (2017).
36. Källénus, G., Correia-Neves, M., Buteme, H., Hamasur, B. & Svenson, S. B. Lipoarabinomannan, and its related glycolipids, induce divergent and opposing immune responses to *Mycobacterium tuberculosis* depending on structural diversity and experimental variations. *Tuberculosis (Edinb.)* **96**, 120–130 (2016).
37. Afonso-Barroso, A. *et al.* Lipoarabinomannan mannose caps do not affect mycobacterial virulence or the induction of protective immunity in experimental animal models of infection and have minimal impact on *in vitro* inflammatory responses. *Cell. Microbiol.* **15**, 660–674 (2013).
38. Briken, V., Porcelli, S. A., Besra, G. S. & Kremer, L. Mycobacterial lipoarabinomannan and related lipoglycans: from biogenesis to modulation of the immune response. *Mol. Microbiol.* **53**, 391–403 (2004).
39. Chatterjee, D. & Khoo, K.-H. Mycobacterial lipoarabinomannan: an extraordinary lipoheteroglycan with profound physiological effects. *Glycobiology* **8**, 113–120 (1998).
40. Dao, D. N. *et al.* *Mycobacterium tuberculosis* lipomannan induces apoptosis and interleukin-12 production in macrophages. *Infect. Immun.* **72**, 2067–2074 (2004).
41. Doz, E. *et al.* Acylation determines the toll-like receptor (TLR)-dependent positive versus TLR2-, mannose receptor-, and SIGNR1-independent negative regulation of pro-inflammatory cytokines by mycobacterial lipomannan. *J. Biol. Chem.* **282**, 26014–26025 (2007).
42. Fukuda, T. *et al.* Critical roles for lipomannan and lipoarabinomannan in cell wall integrity of mycobacteria and pathogenesis of tuberculosis. *MBio* **4**, e00472–e12 (2013).
43. Ishikawa, E., Mori, D. & Yamasaki, S. Recognition of mycobacterial lipids by immune receptors. *Trends Immunol.* **38**, 66–76 (2017).
44. Bomar, L., Brugger, S. D., Yost, B. H., Davies, S. S. & Lemon, K. P. *Corynebacterium accolens* releases antipneumococcal free fatty acids from human nostril and skin surface triacylglycerols. *MBio* **7**, e01725–e15 (2016).
45. Naik, S. *et al.* Commensal-dendritic-cell interaction specifies a unique protective skin immune signature. *Nature* **520**, 104–108 (2015).
46. Linehan, J. L. *et al.* Non-classical immunity controls microbiota impact on skin immunity and tissue repair. *Cell* (in the press).
47. Clark, R. A. *et al.* The vast majority of CLA⁺ T cells are resident in normal skin. *J. Immunol.* **176**, 4431–4439 (2006).
48. Metz, D. *et al.* Immunohistochemical demonstration of immunoglobulin A in human sebaceous and sweat glands. *J. Invest. Dermatol.* **92**, 13–17 (1989).
49. Okada, T., Konishi, H., Ito, M., Nagura, H. & Asai, J. Identification of secretory immunoglobulin A in human sweat and sweat glands. *J. Invest. Dermatol.* **90**, 648–651 (1988).
This study used immunohistochemistry to show that secretory IgA was associated with human sweat glands, and was probably being actively transported in a way similar to the intestine. This study raises the question of how IgA on the skin influences microbiota composition and whether commensal microbes stimulate IgA secretion similarly to gut commensal flora.
50. Fagaras, S. *et al.* Critical roles of activation-induced cytidine deaminase in the homeostasis of gut flora. *Science* **298**, 1424–1427 (2002).
51. Macpherson, A. J., Hunziker, L., McCoy, K. & Lamarre, A. IgA responses in the intestinal mucosa against pathogenic and non-pathogenic microorganisms. *Microbes Infect.* **3**, 1021–1035 (2001).
52. Kawamoto, S. *et al.* The inhibitory receptor PD-1 regulates IgA selection and bacterial composition in the gut. *Science* **336**, 485–489 (2012).
53. van der Waaij, L. A., Limburg, P. C., Mesander, G. & van der Waaij, D. *In vivo* IgA coating of anaerobic bacteria in human faeces. *Gut* **38**, 348–354 (1996).
54. Shroff, K. E., Meslin, K. & Cebra, J. J. Commensal enteric bacteria engender a self-limiting humoral mucosal immune response while permanently colonizing the gut. *Infect. Immun.* **63**, 3904–3913 (1995).
55. Macpherson, A. J. & Uhr, T. Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria. *Science* **303**, 1662–1665 (2004).
56. Vossenkaemper, A. *et al.* A role for gut-associated lymphoid tissue in shaping the human B cell repertoire. *J. Exp. Med.* **210**, 1665–1674 (2013).
57. O’Riordan, K. & Lee, J. C. *Staphylococcus aureus* capsular polysaccharides. *Clin. Microbiol. Rev.* **17**, 218–234 (2004).
58. Cheng, B. L. *et al.* Evaluation of serotypes 5 and 8 capsular polysaccharides in protection against *Staphylococcus aureus* in murine models of infection. *Hum. Vaccin. Immunother.* **13**, 1609–1614 (2017).
59. Zimmermann, M. & Fischbach, M. A. A family of pyrazinone natural products from a conserved nonribosomal peptide synthetase in *Staphylococcus aureus*. *Chem. Biol.* **17**, 925–930 (2010).
60. Wyatt, M. A. *et al.* *Staphylococcus aureus* nonribosomal peptide secondary metabolites regulate virulence. *Science* **329**, 294–296 (2010).
61. Scharschmidt, T. C. *et al.* A wave of regulatory T cells into neonatal skin mediates tolerance to commensal microbes. *Immunity* **43**, 1011–1021 (2015).
62. Scharschmidt, T. C. *et al.* Commensal microbes and hair follicle morphogenesis coordinately drive Treg migration into neonatal skin. *Cell Host Microbe* **21**, 467–477.e5 (2017).
63. Lai, Y. *et al.* Commensal bacteria regulate Toll-like receptor 3-dependent inflammation after skin injury. *Nat. Med.* **15**, 1377–1382 (2009).
64. Loesche, M. *et al.* Temporal stability in chronic wound microbiota is associated with poor healing. *J. Invest. Dermatol.* **137**, 237–244 (2017).
65. Kalan, L. *et al.* Redefining the chronic-wound microbiome: fungal communities are prevalent, dynamic, and associated with delayed healing. *MBio* **7**, e01058–e16 (2016).
66. Feingold, K. R. The outer frontier: the importance of lipid metabolism in the skin. *J. Lipid Res.* **50** (Suppl), S417–S422 (2009).
67. Brandner, J. M. Importance of tight junctions in relation to skin barrier function. *Curr. Probl. Dermatol.* **49**, 27–37 (2016).
68. Natsuga, K. Epidermal barriers. *Cold Spring Harb. Perspect. Med.* **4**, a018218 (2014).
69. McLean, W. H. I. Filaggrin failure—from ichthyosis vulgaris to atopic eczema and beyond. *Br. J. Dermatol.* **175** (Suppl 2), 4–7 (2016).
70. Madison, K. C. Barrier function of the skin: “la raison d’être” of the epidermis. *J. Invest. Dermatol.* **121**, 231–241 (2003).
71. Cleaver, J. E. Common pathways for ultraviolet skin carcinogenesis in the repair and replication defective groups of xeroderma pigmentosum. *J. Dermatol. Sci.* **23**, 1–11 (2000).
72. Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
73. Has, C. & Bruckner-Tuderman, L. The genetics of skin fragility. *Annu. Rev. Genomics Hum. Genet.* **15**, 245–268 (2014).

74. Capell, B. C., Tloughan, B. E. & Orlow, S. J. From the rarest to the most common: insights from progeroid syndromes into skin cancer and aging. *J. Invest. Dermatol.* **129**, 2340–2350 (2009).
75. Tótté, J. E. *et al.* Prevalence and odds of *Staphylococcus aureus* carriage in atopic dermatitis: a systematic review and meta-analysis. *Br. J. Dermatol.* **175**, 687–695 (2016).
76. Tótté, J. E. *et al.* A systematic review and meta-analysis on *Staphylococcus aureus* carriage in psoriasis, acne and rosacea. *Eur. J. Clin. Microbiol. Infect. Dis.* **35**, 1069–1077 (2016).
77. Huang, J. T., Abrams, M., Tloughan, B., Rademaker, A. & Paller, A. S. Treatment of *Staphylococcus aureus* colonization in atopic dermatitis decreases disease severity. *Pediatrics* **123**, e808–e814 (2009).
78. Kobayashi, T. *et al.* Dysbiosis and *Staphylococcus aureus* colonization drives inflammation in atopic dermatitis. *Immunity* **42**, 756–766 (2015).
This study demonstrated potential mechanistic links between dysbiotic skin flora and inflammation in atopic dermatitis by using a mouse model of eczema with ADAM17 deficiency that recapitulates spontaneous development of dysbiotic flora and skin inflammation.
79. Leyden, J. J., Marples, R. R. & Kligman, A. M. *Staphylococcus aureus* in the lesions of atopic dermatitis. *Br. J. Dermatol.* **90**, 525–530 (1974).
This was one of the first studies to demonstrate abundant *Staphylococcus aureus* colonization of patients with atopic dermatitis, even in areas of normal-appearing skin, and established the concept that colonizing microbes can have pathogenic effects without overt infection.
80. Conti, F. *et al.* Association between *Staphylococcus aureus* nasal carriage and disease phenotype in patients affected by systemic lupus erythematosus. *Arthritis Res. Ther.* **18**, 177 (2016).
81. Nakagawa, S. *et al.* *Staphylococcus aureus* virulent PSM α peptides induce keratinocyte alarmin release to orchestrate IL-17-dependent skin inflammation. *Cell Host Microbe* **22**, 667–677.e5 (2017).
82. Liu, H. *et al.* *Staphylococcus aureus* epicutaneous exposure drives skin inflammation via IL-36-mediated T cell responses. *Cell Host Microbe* **22**, 653–666.e5 (2017).
83. Otto, M. Basis of virulence in community-associated methicillin-resistant *Staphylococcus aureus*. *Annu. Rev. Microbiol.* **64**, 143–162 (2010).
84. de Haas, C. J. C. *et al.* Chemotaxis inhibitory protein of *Staphylococcus aureus*, a bacterial antiinflammatory agent. *J. Exp. Med.* **199**, 687–695 (2004).
85. Luong, T. T. & Lee, C. Y. Overproduction of type 8 capsular polysaccharide augments *Staphylococcus aureus* virulence. *Infect. Immun.* **70**, 3389–3395 (2002).
86. Uhlén, M. *et al.* Complete sequence of the staphylococcal gene encoding protein A. A gene evolved through multiple duplications. *J. Biol. Chem.* **259**, 1695–1702 (1984).
87. Palmqvist, N., Patti, J. M., Tarkowski, A. & Josefsson, E. Expression of staphylococcal clumping factor A impedes macrophage phagocytosis. *Microbes Infect.* **6**, 188–195 (2004).
88. Peschel, A. *et al.* Inactivation of the *dlt* operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other antimicrobial peptides. *J. Biol. Chem.* **274**, 8405–8410 (1999).
89. Foster, T. J. Immune evasion by staphylococci. *Nat. Rev. Microbiol.* **3**, 948–958 (2005).
90. Rooijakkers, S. H. M. *et al.* Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. *Nat. Immunol.* **6**, 920–927 (2005).
91. Sonesson, A. *et al.* Identification of bacterial biofilm and the *Staphylococcus aureus* derived protease, staphopain, on the skin surface of patients with atopic dermatitis. *Sci. Rep.* **7**, 8689 (2017).
92. Zhang, L. J. *et al.* Innate immunity. Dermal adipocytes protect against invasive *Staphylococcus aureus* skin infection. *Science* **347**, 67–71 (2015).
This study showed that adipogenesis and adipocyte production of AMPs help to protect against *Staphylococcus aureus* infection via intradermal injection, demonstrating that in addition to keratinocytes and sebocytes, subcutaneous tissues can participate in the immune response to microbes.
93. Joshi, G. S., Spontak, J. S., Klapper, D. G. & Richardson, A. R. Arginine catabolic mobile element encoded *speG* abrogates the unique hypersensitivity of *Staphylococcus aureus* to exogenous polyamines. *Mol. Microbiol.* **82**, 9–20 (2011).
94. Thurlow, L. R. *et al.* Functional modularity of the arginine catabolic mobile element contributes to the success of USA300 methicillin-resistant *Staphylococcus aureus*. *Cell Host Microbe* **13**, 100–107 (2013).
95. Wentworth, A. B., Drage, L. A., Wengenack, N. L., Wilson, J. W. & Lohse, C. M. Increased incidence of cutaneous nontuberculous mycobacterial infection, 1980 to 2009: a population-based study. *Mayo Clin. Proc.* **88**, 38–45 (2013).
96. Merritt, R. W. *et al.* Ecology and transmission of Buruli ulcer disease: a systematic review. *PLoS Negl. Trop. Dis.* **4**, e911 (2010).
97. Houben, R. M. G. J. & Dodd, P. J. The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. *PLoS Med.* **13**, e1002152 (2016).
98. World Health Organization. *Global Tuberculosis Report 2016*; <http://apps.who.int/iris/bitstream/10665/250441/1/9789241565394-eng.pdf?ua=1> (2016).
99. Haley, C. A. Treatment of latent tuberculosis infection. *Microbiol. Spectr.* **5**, TNM17-0039–2016 (2017).
100. Sehgal, V. N. Leprosy. *Dermatol. Clin.* **12**, 629–644 (1994).
101. Talhari, C., Talhari, S. & Penna, G. O. Clinical aspects of leprosy. *Clin. Dermatol.* **33**, 26–37 (2015).
102. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1151–1210 (2017).
103. Wansbrough-Jones, M. & Phillips, R. Buruli ulcer: emerging from obscurity. *Lancet* **367**, 1849–1858 (2006).
104. Marion, E. *et al.* Mycobacterial toxin induces analgesia in Buruli ulcer by targeting the angiotensin pathways. *Cell* **157**, 1565–1576 (2014).
This study shows that mycolactone, a virulence factor produced by the cutaneous pathogen *Mycobacterium ulcerans*, causes analgesia by directly binding to the angiotensin II receptor on nerve cells and triggering downstream potassium channel activation and resultant cell hyperpolarization.
105. Kashem, S. W. *et al.* Nociceptive sensory fibers drive interleukin-23 production from CD301b⁺ dermal dendritic cells and drive protective cutaneous immunity. *Immunity* **43**, 515–526 (2015).
This study shows that neurons can participate directly in the immune response to microbes. Cutaneous sensory neurons are directly activated by *Candida albicans* and subsequently stimulate dermal dendritic cells to produce IL-23, thus driving protective immunity by IL-17A-producing dermal T cells.
106. Chiu, I. M. *et al.* Bacteria activate sensory neurons that modulate pain and inflammation. *Nature* **501**, 52–57 (2013).
107. Cardoso, V. *et al.* Neuronal regulation of type 2 innate lymphoid cells via neuromedin U. *Nature* **549**, 277–281 (2017).
108. Bellono, N. W. *et al.* Enterochromaffin cells are gut chemosensors that couple to sensory neural pathways. *Cell* **170**, 185–198 (2017).
109. Fung, T. C. *et al.* Lymphoid tissue-resident commensal bacteria promote members of the IL-10 cytokine family to establish mutualism. *Immunity* **44**, 634–646 (2016).
110. Sonnenberg, G. F. *et al.* Innate lymphoid cells promote anatomical containment of lymphoid-resident commensal bacteria. *Science* **336**, 1321–1325 (2012).
111. Davis, J. M. & Ramakrishnan, L. The role of the granuloma in expansion and dissemination of early tuberculous infection. *Cell* **136**, 37–49 (2009).
112. Horsburgh, C. R. J., Jr. Priorities for the treatment of latent tuberculosis infection in the United States. *N. Engl. J. Med.* **350**, 2060–2067 (2004).
113. Adams, K. N. *et al.* Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. *Cell* **145**, 39–53 (2011).
114. Schnappinger, D. *et al.* Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. *J. Exp. Med.* **198**, 693–704 (2003).
115. Cunningham, A. F. & Spreadbury, C. L. Mycobacterial stationary phase induced by low oxygen tension: cell wall thickening and localization of the 16-kilodalton α -crystallin homolog. *J. Bacteriol.* **180**, 801–808 (1998).
116. Rittershaus, E. S. C., Baek, S.-H. & Sasseti, C. M. The normalcy of dormancy: common themes in microbial quiescence. *Cell Host Microbe* **13**, 643–651 (2013).
117. Bartek, I. L. *et al.* *Mycobacterium tuberculosis* Lsr2 is a global transcriptional regulator required for adaptation to changing oxygen levels and virulence. *MBio* **5**, e01106–e01114 (2014).
118. Eoh, H. *et al.* Metabolic anticipation in *Mycobacterium tuberculosis*. *Nat. Microbiol.* **2**, 201784 (2017).
119. Galagan, J. E. *et al.* The *Mycobacterium tuberculosis* regulatory network and hypoxia. *Nature* **499**, 178–183 (2013).
120. Lin, P. L. *et al.* Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. *Nat. Med.* **20**, 75–79 (2014).
121. Hannigan, G. D. *et al.* The human skin double-stranded DNA virome: topographical and temporal diversity, genetic enrichment, and dynamic associations with the host microbiome. *MBio* **6**, e01578–e15 (2015).
122. Hickman, H. D. *et al.* Anatomically restricted synergistic antiviral activities of innate and adaptive immune cells in the skin. *Cell Host Microbe* **13**, 155–168 (2013).
This study used intravital multiphoton microscopy to demonstrate that effector CD8⁺ T cells respond to cutaneous vaccinia virus infection by killing infected monocytes in the periphery but not infected keratinocytes in the center, thus highlighting the spatial complexity and specificity of immune cell dynamics in the skin.
123. Cush, S. S. *et al.* Locally produced IL-10 limits cutaneous vaccinia virus spread. *PLoS Pathog.* **12**, e1005493 (2016).
124. Carbone, F. R. Tissue-resident memory T cells and fixed immune surveillance in nonlymphoid organs. *J. Immunol.* **195**, 17–22 (2015).
125. Mueller, S. N., Gebhardt, T., Carbone, F. R. & Heath, W. R. Memory T cell subsets, migration patterns, and tissue residence. *Annu. Rev. Immunol.* **31**, 137–161 (2013).
126. Lanas, A. & Chan, F. K. L. Peptic ulcer disease. *Lancet* **390**, 613–624 (2017).
127. Kalisiperati, P. *et al.* Inflammation, DNA damage, *Helicobacter pylori* and gastric tumorigenesis. *Front. Genet.* **8**, 20 (2017).
128. Peek, R. M., Jr & Blaser, M. J. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nat. Rev. Cancer* **2**, 28–37 (2002).
129. Kienesberger, S. *et al.* Gastric *Helicobacter pylori* infection affects local and distant microbial populations and host responses. *Cell Reports* **14**, 1395–1407 (2016).

130. Lehman, H. Skin manifestations of primary immune deficiency. *Clin. Rev. Allergy Immunol.* **46**, 112–119 (2014).
131. Barnard, E., Shi, B., Kang, D., Craft, N. & Li, H. The balance of metagenomic elements shapes the skin microbiome in acne and health. *Sci. Rep.* **6**, 39491 (2016).
132. Agak, G. W. *et al.* *Propionibacterium acnes* Induces an IL-17 response in acne vulgaris that is regulated by vitamin A and vitamin D. *J. Invest. Dermatol.* **134**, 366–373 (2014).
133. Fitz-Gibbon, S. *et al.* *Propionibacterium acnes* strain populations in the human skin microbiome associated with acne. *J. Invest. Dermatol.* **133**, 2152–2160 (2013).
134. Ring, H. C. *et al.* The follicular skin microbiome in patients with hidradenitis suppurativa and healthy controls. *JAMA Dermatol.* **153**, 897–905 (2017).
135. Wollenberg, M. S. *et al.* *Propionibacterium*-produced coproporphyrin III induces *Staphylococcus aureus* aggregation and biofilm formation. *MBio* **5**, e01286–e14 (2014).
136. Shu, M. *et al.* Fermentation of *Propionibacterium acnes*, a commensal bacterium in the human skin microbiome, as skin probiotics against methicillin-resistant *Staphylococcus aureus*. *PLoS One* **8**, e55380 (2013).
137. Barton, E. S. *et al.* Herpesvirus latency confers symbiotic protection from bacterial infection. *Nature* **447**, 326–329 (2007).
138. Yager, E. J. *et al.* γ -Herpesvirus-induced protection against bacterial infection is transient. *Viral Immunol.* **22**, 67–72 (2009).
139. Perry, S. *et al.* Infection with *Helicobacter pylori* is associated with protection against tuberculosis. *PLoS One* **5**, e8804 (2010).
140. Arnold, I. C. *et al.* *Helicobacter pylori* infection prevents allergic asthma in mouse models through the induction of regulatory T cells. *J. Clin. Invest.* **121**, 3088–3093 (2011).
141. Cheung, G. Y. C. & Otto, M. Understanding the significance of *Staphylococcus epidermidis* bacteremia in babies and children. *Curr. Opin. Infect. Dis.* **23**, 208–216 (2010).
142. Rohlfke, F. & Stollman, N. Fecal microbiota transplantation in relapsing *Clostridium difficile* infection. *Therap. Adv. Gastroenterol.* **5**, 403–420 (2012).
143. Panigrahi, P. *et al.* A randomized synbiotic trial to prevent sepsis among infants in rural India. *Nature* **548**, 407–412 (2017).
This study was the first, to our knowledge, to use an oral synbiotic (*Lactobacillus plantarum* and fructooligosaccharide) to promote effective gut colonization of the inoculated bacterium and reduce neonatal sepsis.
144. Nakatsuji, T. *et al.* Sebum free fatty acids enhance the innate immune defense of human sebocytes by upregulating β -defensin-2 expression. *J. Invest. Dermatol.* **130**, 985–994 (2010).
145. Falchook, G. S. *et al.* Responses of metastatic basal cell and cutaneous squamous cell carcinomas to anti-PD1 monoclonal antibody REGN2810. *J. Immunother. Cancer* **4**, 70 (2016).
146. Morris, V. K. *et al.* Nivolumab for previously treated unresectable metastatic anal cancer (NCI9673): a multicentre, single-arm, phase 2 study. *Lancet Oncol.* **18**, 446–453 (2017).
147. Mahoney, K. M., Freeman, G. J. & McDermott, D. F. The next immune-checkpoint inhibitors: PD-1/PD-L1 blockade in melanoma. *Clin. Ther.* **37**, 764–782 (2015).
148. Cassler, N. M. & Brownell, I. PD-1 checkpoint blockade is an emerging treatment for Merkel cell carcinoma. *Br. J. Dermatol.* **176**, 18 (2017).
149. Sivan, A. *et al.* Commensal *Bifidobacterium* promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* **350**, 1084–1089 (2015).
This study and the next one demonstrated that the gut microbial composition can alter immunotherapies at distant sites, suggesting that microbe-immune interactions at barrier sites can have far-reaching effects at other barrier sites or systemically.
150. Vétizou, M. *et al.* Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science* **350**, 1079–1084 (2015).
151. Golden, J. B. *et al.* Chronic, not acute, skin-specific inflammation promotes thrombosis in psoriasis murine models. *J. Transl. Med.* **13**, 382 (2015).
152. Santilli, S. *et al.* Visualization of atherosclerosis as detected by coronary artery calcium and carotid intima-media thickness reveals significant atherosclerosis in a cross-sectional study of psoriasis patients in a tertiary care center. *J. Transl. Med.* **14**, 217 (2016).
153. Wang, Y. *et al.* Chronic skin-specific inflammation promotes vascular inflammation and thrombosis. *J. Invest. Dermatol.* **132**, 2067–2075 (2012).
This study used a mouse model of psoriasis (keratinocyte-specific Tie2 transgene expression) to mechanistically link skin inflammation to the development of resultant aortic root inflammation and also to show that subsequent treatment of the skin disease can eliminate not only skin inflammation but also systemic vascular inflammation.
154. Evensen, K. *et al.* Increased subclinical atherosclerosis in patients with chronic plaque psoriasis. *Atherosclerosis* **237**, 499–503 (2014).
155. Tomura, M. *et al.* Activated regulatory T cells are the major T cell type emigrating from the skin during a cutaneous immune response in mice. *J. Clin. Invest.* **120**, 883–893 (2010).
156. Tomura, M. *et al.* Tracking and quantification of dendritic cell migration and antigen trafficking between the skin and lymph nodes. *Sci. Rep.* **4**, 6030 (2014).
157. Geherin, S. A. *et al.* The skin, a novel niche for recirculating B cells. *J. Immunol.* **188**, 6027–6035 (2012).
This study was one of the first to demonstrate that B cells actively traffic in and out of the skin, even in uninflamed skin, suggesting that B cells play an active and previously underappreciated role in skin homeostasis.
158. Belkaid, Y. & Segre, J. A. Dialogue between skin microbiota and immunity. *Science* **346**, 954–959 (2014).
159. Gallo, R. L. & Hooper, L. V. Epithelial antimicrobial defence of the skin and intestine. *Nat. Rev. Immunol.* **12**, 503–516 (2012).
160. Nithya, S., Radhika, T. & Jedy, N. Loricrin—an overview. *J. Oral Maxillofac. Pathol.* **19**, 64–68 (2015).
161. Iguchi, A. *et al.* A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res.* **22**, 101–107 (2015).

Acknowledgements We apologize for not having cited all papers relevant to this expanding field of research (in particular, older literature) because of space constraints and editorial limits. This work was supported by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases (Y.B.), DP1 DK113598 (M.A.F.), R01 DK110174 (M.A.F.), an HHMI-Simons Faculty Scholars Award (M.A.F.), a Fellowship for Science and Engineering from the David and Lucile Packard Foundation (M.A.F.), a Burroughs Wellcome Investigators in the Pathogenesis of Infectious Disease Award (M.A.F.) and the Dermatology Foundation (Y.E.C.).

Author Contributions Y.E.C., M.A.F. and Y.B. conceptualized the article structure, content, and figures, and wrote and edited the manuscript and figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.A.F. (fischbach@fischbachgroup.org) or Y.B. (YBelkaid@niaid.nih.gov).

Reviewer Information *Nature* thanks T. Scharschmidt and the other anonymous reviewer(s) for their contribution to the peer review of this work.

The rise of three-dimensional human brain cultures

Sergiu P. Pașca¹

Pluripotent stem cells show a remarkable ability to self-organize and differentiate *in vitro* in three-dimensional aggregates, known as organoids or organ spheroids, and to recapitulate aspects of human brain development and function. Region-specific 3D brain cultures can be derived from any individual and assembled to model complex cell–cell interactions and to generate circuits in human brain assembloids. Here I discuss how this approach can be used to understand unique features of the human brain and to gain insights into neuropsychiatric disorders. In addition, I consider the challenges faced by researchers in further improving and developing methods to probe and manipulate patient-derived 3D brain cultures.

Understanding the principles that underlie the assembly of cells into tissues and of tissues into organs is a fundamental goal in biology. Such understanding requires not just observation, but also the ability to construct and deconstruct complex, developing structures. This has been particularly challenging when studying the central nervous system (CNS) in humans, in part because of its complexity, but also because of poor accessibility to all stages of development and lack of functional tissue preparations. In other branches of medicine, such as haematology and oncology, easy access to tissue samples has led to a comprehensive understanding of organ development and substantial therapeutic advances. Therefore, there is a pressing need to develop functional, realistic and personalized models of the developing human brain so that we can better understand its unique biology and gain mechanistic insights into neuropsychiatric disorders.

Several recent conceptual and technological advances are now converging to make human brain tissue more accessible for study. First, the ability to culture pluripotent stem cells, including human embryonic stem (hES) cells, *in vitro*^{1,2}. Second, the possibility to reprogram somatic cells into induced pluripotent stem (iPS) cells³ and subsequently to promote their differentiation into neurons⁴, or to shortcut this process and directly derive neurons^{5,6}. Third, progress in building 3D brain cultures as well as advances in biomaterials, CRISPR–Cas9-based genome engineering⁷ and highly-parallel single-cell transcriptomics⁸. Combined, these advances open opportunities for understanding the assembly of the human brain and how this may go awry in disease. This review discusses advances in building 3D human brain cultures, such as neural organoids or spheroids, and describes how these cultures may help researchers capture normal and abnormal organogenesis *in vitro*. While these approaches may bring access to previously inaccessible aspects of human biology, such as developmental processes in late human gestation, they are still models. As George Box pointed out, “all models are wrong but some are useful,”⁹ and their value ultimately resides in their ability to provide testable predictions. Therefore, an important goal of this overview will also be to highlight the advantages and disadvantages of the various approaches for engineering *in vitro* models of the human nervous system.

From pluripotent cells to brain cells in a dish

What principles guide organogenesis? Immanuel Kant astutely described life as a “self-organized, self-reproducing” process¹⁰. Self-organization implies the formation of ordered structures from relatively homogeneous elements in the absence of an external pattern. In embryology, this

involves a dynamic process that starts with a relatively homogenous group of cells that are capable of differentiation and self-patterning and that respond to external forces. The combined action of internal (genetic, biochemical) and external (mechanical) inputs, as well as stochastic events, lead to symmetry breaking, cell rearrangements and non-uniform but controlled spatiotemporal growth. These processes result in emergent properties of the developing structure. For instance, self-assembly involves a rearrangement of elements. The concept of self-assembly originated in chemistry, as seen in Rayleigh–Bernard convection, but it has been extensively described in living organisms. Single dissociated cells obtained from amphibians will meaningfully self-sort when pH is restored¹¹ (Fig. 1a). Similarly, a single hydra can be dissociated into single cells, which then reassemble to recreate the entire organism¹². Cell arrangement mediated by surface proteins is, however, not the only mechanism for self-assembly. For example, periodic waves of gene expression, which can be synchronized across groups of cells, have been shown to participate in the self-assembly of dissociated cells from the presomatic mesoderm¹³.

Neural differentiation of pluripotent cells

Human organogenesis follows many of the same developmental patterns seen in other species. The nervous system develops from a single tube that undergoes disproportionate enlargement of the anterior side (Fig. 1b). This process of generating biological tissue shape, also known as morphogenesis, involves local proliferation and patterning, complex cell–cell interactions, cell fate specification and long-distance migration. For instance, the formation of the cerebral cortex in humans involves massive proliferation of progenitors in various domains located close to the ventricle, followed by the orderly generation and arrangement of glutamatergic neurons, starting with lower layers and finishing with upper layer neurons positioned close to the pia¹⁴. Other cells, such as GABAergic neurons, migrate into the cerebral cortex after being specified in distant regions, and corticogenesis continues postnatally with the generation of glial cells. In other parts of the nervous system, the final arrangement can involve *en masse* physical movements of cells, such as in the case of the evagination and invagination that underlie optic cup formation.

Methods for inducing the differentiation of mouse and human pluripotent stem (mPS and hPS) cells *in vitro* can, surprisingly, recapitulate some of these elaborate processes even in 2D cultures¹⁵. Colonies of hPS cells can be micropatterned to recapitulate gastrulation-like events¹⁶. Contrary to expectations, this phenomenon occurs without cell motility

¹Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, USA.

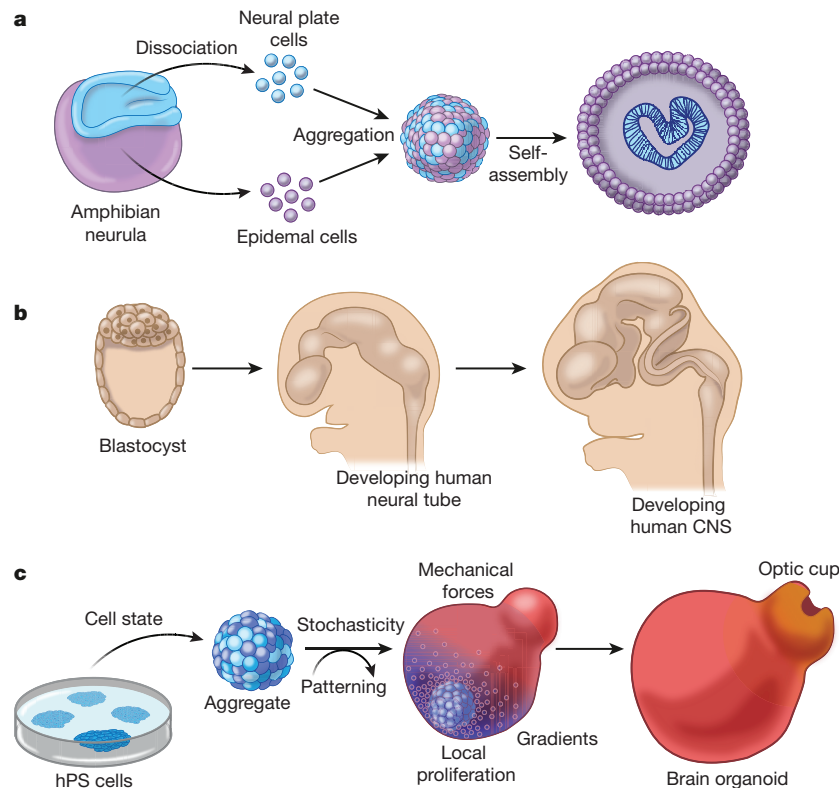


Figure 1 | Self-organization and organogenesis. **a**, Mixed single cells dissociated from the neural plate or the epidermis of an amphibian neurula can self-assemble and generate epidermis-like layers around a neural tube-like structure. **b**, Development of a complex CNS from a neural tube derived from hPS cells isolated from the inner mass of a

blastocyst. **c**, Self-organization of brain organoids from hPS cells depends on cell state (for example, 'naïve' versus 'primed'), size of initial aggregate, self-patterning or external patterning with growth factors and small molecules, local proliferation, mechanical forces and stochastic factors.

and may be the result of a Turing 'reaction–diffusion' mechanism based on signalling molecules¹⁷ or an 'edge-sensing' mechanism in which cells respond differentially according to their position within the colony¹⁸. Early efforts to derive neural cells from pluripotent stem cells *in vitro* triggered differentiation by giving cells more degrees of freedom as small aggregates called embryoid bodies⁴ and leveraged the fact that the predominant germ layer fate is ectodermal, even in *Xenopus*¹⁹. Subsequently, double inhibition of the SMAD pathway in hPS cells grown in high density 2D cultures was shown to be sufficient to generate a high proportion of neural precursors²⁰. Neuroepithelial cells display a high degree of polarity and form neural rosettes around a pseudo-lumen²¹. This is a common pattern of neural cell organization, which has also been observed in CNS tumours and the ectodermal component of teratomas. The default fate of these neural precursors is forebrain, and minimal intervention is required to recapitulate the sequential generation of layer-specific cortical neurons in both rodent and human 2D cultures^{22–25}. This anterior or rostral default state can be overturned, and cells can be converted to more caudal fates by patterning with small molecules and growth factors to generate midbrain, striatal or spinal cord neurons^{26,27}. Moreover, stromal co-culture of mPS cells can induce the formation of 3D structures that include diverse cell types, such as eye-related structures²⁸.

However, there are limitations to stem cell differentiation in 2D cultures. Interactions with plastic surfaces prevail over interactions between cells or between cells and the extracellular matrix (ECM). Gradients of patterning molecules, gases and nutrients are dispersed and the interactions of growth factors with heparin sulfate or proteoglycans are altered. Apical–basal polarity is changed and migration is not constrained; at low density, neural progenitors move in a slow, amoeba-like way, but at high density they glide rapidly²⁹. The stiffness of plastic dishes is not physiological and many cells isolated from organs or tumours become flat when cultured in 2D, altering their proliferation

rate and differentiation status. Not surprisingly, drug screening trials in 2D cultures yield different results from those carried out in 3D cultures³⁰. These issues have prompted the development of culture systems that recapitulate more complex cell–cell interactions and cell diversity, mature to later stages in development and that show higher levels of functionality (see Box). Because of the intention to more closely model the cyto-architecture of organs, these 3D cultures are referred to as organoids or organ spheroids.

Starting with a small number of pluripotent cells, 3D cultures rely upon genetically encoded self-organization³¹ to generate *in vitro*, and without an existing pattern, polarized floating structures that resemble *in vivo* tissue (Fig. 1c). As with all nonlinear systems, the state of initial elements is fundamental. For example, seeding density can influence fate choices. Early heterogeneity in cell states ('naïve' or ground state versus 'primed'), stochastic processes and cell–cell interactions in aggregates contribute to symmetry breaking early on in 3D ensembles. Exogenous molecules or physical confinement of these aggregates, local differentiation and subsequent secretion of patterning molecules give rise to molecular gradients and reaction–diffusion phenomena. These processes in turn can cause localized proliferation and changes in mechanical forces, further leading to specialization and reassembly of cells.

The generation of an optic cup is probably the best way to illustrate the surprising capacity of pluripotent stem cells to self-organize in 3D cultures^{32,33}. In this case, modulation of the Wnt pathway is used to develop retinal epithelium and vesicle-like structures. These structures show local mechanical autonomy, with proliferation and cytoskeletal changes in cells at key locations resulting in spontaneous curving and formation of the optic cup. Self-organizing phenomena and even multi-germ layer lineages have also been observed in cultures of non-brain tissues, such as kidney³⁴ and lung³⁵. Moreover, gastrointestinal-related 3D cultures that include crypt-like structures can be derived from single stem cells isolated from primary tissue³⁶.

Developments in 3D neural differentiation

What are the approaches for deriving 3D brain cultures, and what aspects can be recapitulated with this platform? Even the earliest reports of neural cultures describe efforts to maintain intact tissue architecture *in vitro*. As early as 1907, Harrison took frog neural tube and established hanging drop cultures by attaching tissue fragments to a glass coverslip in coagulated serum or lymph, so that “growing nerves could be brought under direct observation while alive”³⁷. This 3D preparation could be maintained for weeks *in vitro* and was helpful in culturing the poliovirus in spinal cord cells from monkeys. Later, roller tubes and semipermeable membranes were used to grow slices of brain tissue in organotypic cultures³⁸. This system maintains some of the 3D architecture and connectivity of the source tissue and can be subsequently grafted *in vivo* for vascularization (for example, into the lateral angle of the eye³⁹). It can also be used to derive tissue explants, in which groups of cells or slices from different brain regions are kept in close juxtaposition to allow specific cell–cell interactions. But to grow neural stem cells that produce various CNS cell lineages, 3D aggregation into structures called neurospheres is essential⁴⁰. The ability to isolate neural stem cells and to differentiate hPS cells in 2D prompted several approaches for deriving 3D brain tissue (Fig. 2).

One direction has been to build upon culture chamber systems, which have proved useful in identifying growth factors by cellular and subcellular isolation. This top-down approach, known as organ-on-a-chip, uses physical channels to position cell types, create gradients and control the flow of nutrients, and to provide spatial and temporal control of the cellular environment. For instance, microchip models of the blood–brain barrier include endothelial cells on one side of a membrane and neurons, pericytes and astrocytes on the other side, and can test the effects of cytokines⁴¹. This system uses reverse-engineering principles and provides rigorous control of variables, but depends on detailed knowledge of the organ and its physiology.

An alternative approach has been to rely on spontaneous morphogenesis in cell aggregates, such as organoids or organ spheroids. Adult stem cells or cells differentiated from pluripotent stem cells as well as tumour cells, can be used to derive organoids in suspension or embedded into extracellular matrices. This has been elegantly demonstrated by the Clevers laboratory for the gastrointestinal tract^{36,42–44}. Organoid approaches allow more degrees of freedom in long-term cultures that give rise to cell diversity, complex cell–cell interactions and unique physical structures. When starting with hPS cells, there are two ways of differentiating organoids: undirected and directed (Fig. 2).

In directed differentiation approaches, aggregates of hPS cells are instructed to acquire an ectodermal fate and subsequently specified to become region-specific organoids or organ spheroids. The pioneering methods developed by the late Yoshiki Sasai involved 3D aggregation of mPS⁴⁵ or hPS⁴⁶ cells and culture in U- or V-bottomed wells followed by 2D plating of differentiated cells at a later stage. These experiments showed that the size of the initial clusters and the use of small molecules for survival were essential, and that lineage reporters and mechanical dissection can enrich the cultures for specific brain regions. An alternative approach used neural specification in 2D followed by 3D cultures of rosettes, yielding a combination of dorsal and ventral forebrain and maturation up to the first trimester stage of brain development⁴⁷. Our group introduced a simple method for deriving dorsal forebrain in 3D that involves lifting intact colonies of human iPS cells, followed by neuralization and culture exclusively in suspension, without an extracellular matrix or culture in a bioreactor⁴⁸. These spherical cultures grow up to 4 mm in diameter, contain equal proportions of deep and superficial layer cortical neurons as well as non-reactive astrocytes, and after approximately 9–10 months mature to resemble postnatal stages⁴⁹. Similarly, the Song and Ming groups used patterning and miniaturized spinning bioreactors to obtain forebrain organoids and to derive midbrain or hypothalamus organoids⁵⁰.

In undirected organoid differentiation, such as the techniques developed by the Knoblich group, hPS cells are suspended and grown in an

BOX 1

2D or 3D—that is the question

With many available options and the excitement surrounding 3D culture techniques, deciding what differentiation approach to use may not be easy. Both 2D and 3D neural differentiation methods have advantages and disadvantages for answering different questions.

Two-dimensional neural cultures can be used to study the neural stem cells and disease mechanisms underlying defects in neural progenitors. For instance, Iefremova *et al.*⁶⁰ first identified defects in organoids derived from individuals with Miller–Dieker syndrome, but to dissect the mechanism, they switched to a 2D culture system to find alterations in N-cadherin– β -catenin–Wnt signalling in radial glia. The scalability of 2D neural cultures make this system more useful for large-scale drug testing or for genome-wide CRISPR–Cas9 screens. Imaging assays and some morphological studies (for example, of dendrite complexity) are also easier to implement in 2D. Directed monolayer differentiation approaches can also provide the high-purity cultures necessary for therapeutic transplantation studies.

Three-dimensional neural cultures can be used over long periods—for almost two years⁴⁹—and provide access to a large diversity of cell types and functional maturation states. The cytoarchitecture and cell–cell interactions are reminiscent of *in vivo* neural tissue. The cross-talk between specific cell types, such as astrocytes and neurons or oligodendrocytes, in the context of synaptogenesis or myelination, may be more informative in a 3D setting. Certain cellular phenotypes can best be studied in 3D. For instance, modular brain assembloids can be used to model inter-regional communication and dissect cell-autonomous versus non-autonomous effects⁶⁸. Cortical interneurons display minimal migration in 2D systems, but accurately recapitulate saltatory movements in 3D cultures when compared to fetal tissue⁶⁸.

extracellular matrix, such as Matrigel, in spinning bioreactors^{51,52}. Owing to the lack of inductive signals, these 3D cultures exhibit a variety of brain region identities and non-neural fates. Single-cell transcriptomic studies in undirected organoids^{53,54} confirmed that dorsal and ventral forebrain cells are mixed with cells from other brain regions, such as retina, hind-brain and midbrain, and co-exist with choroid plexus and mesodermal cells. Recent work showed that individual organoids can acquire different fates and demonstrated the presence of various cell classes found in the mouse retina⁵⁴. These differentiation techniques have a higher degree of stochasticity than directed differentiation, and early conditions could have large effects. Unconstrained organization leads to unique morphologies and levels of maturation, and the challenges in identifying and removing certain populations of cells can lead to non-physiological cell–cell interactions. Variability in undirected organoids may be related to inconsistency in neural induction⁵⁵, and the tendency has been to constrain cell fates using small molecules^{56,57} or fibre microfilaments⁵⁵. On the other hand, the high degree of diversity in these cultures may allow researchers to explore human CNS diversity and to map disease genes onto specific cell types.

Several directed approaches have been used to derive CNS regions in 3D cultures. Differentiation of hES cells in 40% oxygen and with up to 2% Matrigel dissolved in the medium⁵⁸ can yield forebrain cultures that show rolling and the formation of curvature with rostro-caudal polarization. This approach also generates ventral forebrain regions as well as abundant choroid plexus⁵⁹, but mostly dorsal forebrain when oxygen is removed in a subsequent modification⁶⁰. By manipulating Wnt and bone morphogenetic protein (BMP) signalling, the fate of these cultures can be shifted more medially to derive hippocampus-like 3D cultures with both granule and pyramidal neurons⁶¹. Alternatively,

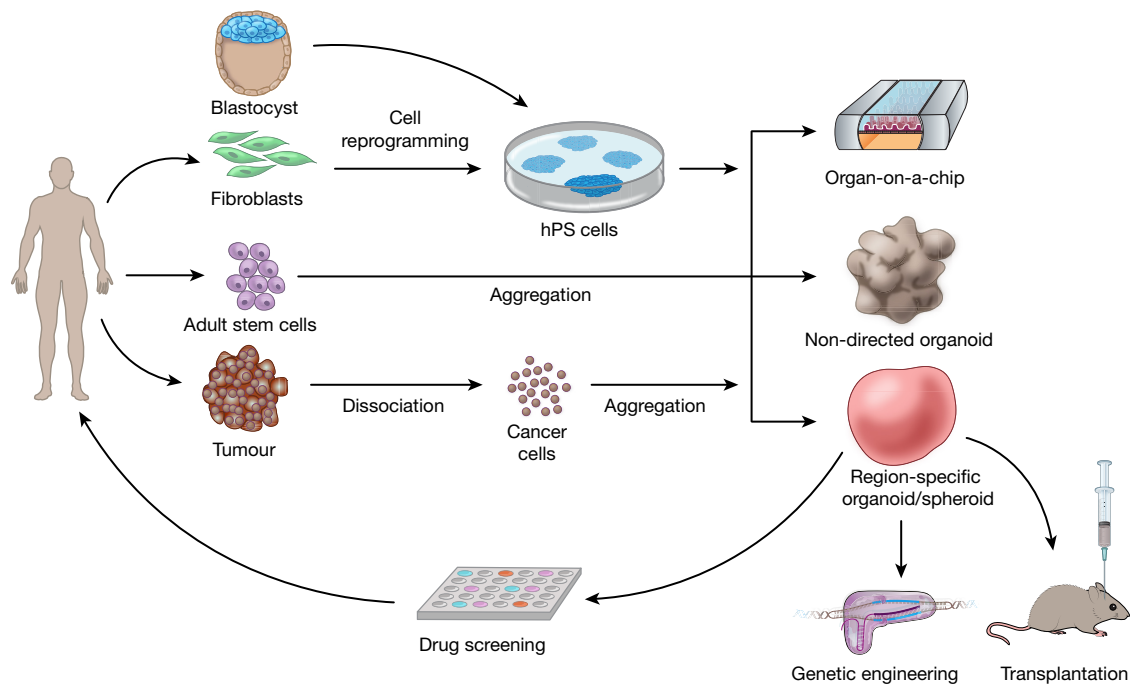


Figure 2 | Different approaches for deriving human brain 3D cultures. hPS cells derived from a blastocyst or by reprogramming of somatic cells, adult stem cells or cancer cells derived from primary tissue can be used to derive microfluidics-based organs-on-a-chip (top), undirected organoids

(middle), and region-specific brain organoids or organ spheroids (bottom). These 3D cultures can be manipulated with CRISPR–Cas9 genome-editing technologies, transplanted into animals or used for drug screening.

growing hPS cells at high density in spinner flasks can yield large numbers of motor neurons⁶².

Some of the region-specific organoid approaches result in unique features, such as the presence in of neuromelanin in midbrain organoids⁶³, an insoluble, dark polymer that becomes apparent postnatally in the substantia nigra^{64,65}. Another unique aspect is the ability to generate distinct niches in the same preparation. Large aggregates of mES cells in the presence of sonic hedgehog (Shh) can generate oral ectoderm on the surface that can invaginate and contact a hypothalamic primordium present within, thus mimicking the development of the adenohypophysis⁶⁶. In the presence of high fibroblast growth factor 2 (FGF2), insulin and FGF19, cerebellum-like organoids develop an elongated, polarized cerebellar plate that generates precursors for GABA (γ -aminobutyric acid)-releasing Purkinje cells and, at the edge, a rhombic-lip region containing granule progenitors⁶⁷. However, in these cerebellum organoids as well as in forebrain organoids that include multiple domains (ventral, dorsal, medial), interactions between these regions are spatially unpredictable and asking specific development- or disease-related questions has been challenging. To address this issue, we introduced controlled assembly of 3D brain cultures⁶⁸ in what I will refer to as brain assembloids, to direct and probe more complex cell–cell interactions and to generate, as in electrical engineering, circuits from parts (Fig. 3).

Developing human brain assembloids

It has been particularly challenging to study cell migration, inter-regional interactions and circuit assembly in the human CNS because it is not possible to obtain intact tissue at later stages of *in utero* development. For instance, the formation of cortical circuits involves not just connectivity between layers of glutamatergic excitatory neurons, but also the integration of around 20% GABAergic interneurons^{69,70}. Interestingly, these interneurons are generated not in the dorsal forebrain, like glutamatergic cells, but in the ventral forebrain (subpallium)^{69,71–73} and must migrate for months over long distances, beginning at mid-fetal human development⁷². Dysfunctional cross-talk between these two cortical cell types is thought to contribute to the pathophysiology of several neuropsychiatric disorders, including epilepsy and autism spectrum disorders (ASD)^{74,75}, but high-resolution probing and manipulation of cortical ensembles in

humans has not been possible. To address this problem, we specified subdomains of the forebrain that functionally interact in development (Fig. 3a) and generated region-specific organoids resembling either the dorsal pallium or the subpallium, and subsequently fused them⁶⁸. This modular system enabled us to monitor the saltatory migration of interneurons towards the cerebral cortex and to identify phenotypes in patient-derived cells. It also demonstrated that interneurons successfully integrate into a synaptically connected microcircuit. This approach has been subsequently used by other groups to model forebrain interactions^{56,76}.

Another way to build assembloids is by directly mixing cells of different lineages or by adding cells or biomaterials that have organizer-like capabilities. For instance, neural progenitors, endothelial cells, mesenchymal cells and microglia or macrophages have been mixed in a peptide-functionalized hydrogel and then used to test neurotoxicity⁷⁷. Moreover, specific populations of spinal cord neurons derived from mES cells have been incorporated into aggregates to build rhythmically active circuits⁷⁸. Wokman *et al.* built a gut–neural assembloid using neural crest cells and intestinal organoids⁷⁹, in which neural cells migrated into the mesenchyme of the intestinal-like tissue, self-organized and gave rise to rhythmic waves. When transplanted into rodents, these organoids showed electromechanical coupling and propagating contractions. CNS assembloids could also be built to study myelination by the addition or *in situ* generation of oligodendrocytes, especially as *in vitro* myelination methods are currently limited, or to model primary or metastatic brain cancer by the addition of tumour cells or assembly with cancer organoids.

Assembloids have the potential to capture more complex inter-regional brain interactions, building upon models in rodents that utilize spatially positioned brain explants. For instance, in rodent cortico-thalamic explants, only multipolar neurons of the thalamus project towards the cortex; these stop on layer 4 pyramidal neurons, even when placed close to the pia^{80,81}. Vice versa, cortical neurons from deep layers project into the thalamus. The developmental stage for establishing these interactions matters, and the unique cytoarchitecture of the sensory cortex is present only when explants are not cut tangentially⁸². The generation of cortico-thalamic assembloids using fusion or spatio-temporally controlled patterning (Fig. 3b) would allow the study of early thalamic projections

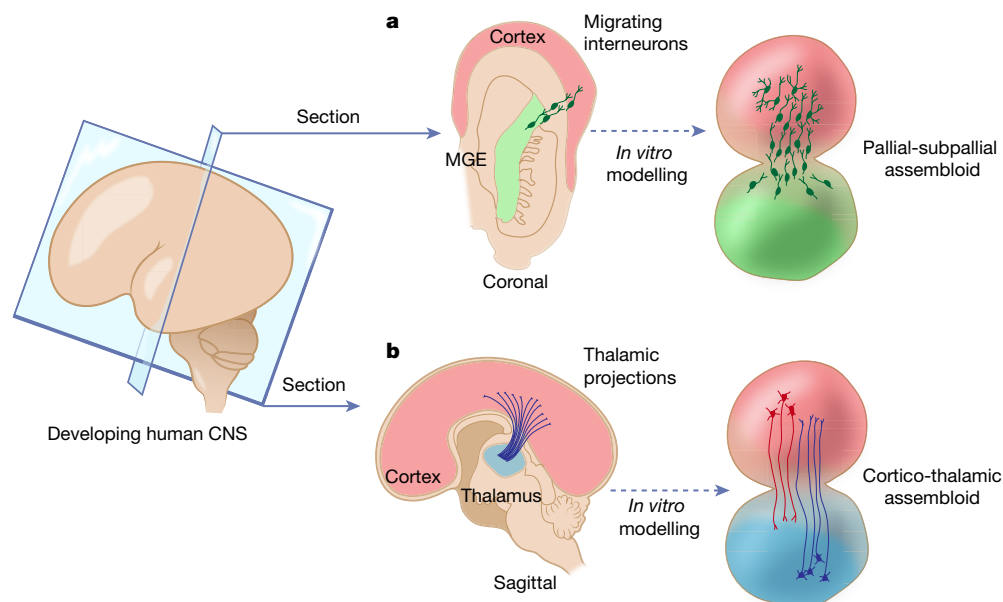


Figure 3 | Human brain assembloids. Cross-sections through the developing human brain at gestational week 17, showing the cerebral cortex (red), the medial ganglionic eminences (MGE, green) and the thalamus (blue). Interneurons from the MGE migrate tangentially to populate the dorsal pallium. Thalamic neurons project to the cortical subplate and then onto layer 4 cortical neurons, while deep layer cortical

neurons project back to the thalamus. **a**, Pallial–subpallial assembloids show modelling of GABAergic interneuron migration and functional integration into cortical circuits. **b**, Cortico–thalamic assembloids illustrate projections from deep layer cortical neurons onto thalamic neurons and projections of thalamic neurons onto layer 4 cortical neurons.

into the human subplate and the role of thalamic activity in building cortical networks. Moreover, it could be used as a platform to investigate patient-derived cultures and verify, for instance, the role of dopamine receptors in mediating thalamocortical dysfunction in individuals with 22q11.2 deletion syndrome⁸³. Similarly, a modular approach to specifying and assembling other circuits could bring insights into dysfunctions of cortico-striatal projections, cortico-spinal tracts, the meso-cortical pathway or cortico-hippocampal projections. Importantly, functional interactions between these cell types may lead to novel features for interrogation *in vitro*, such as the maturation of muscle fibres following innervation, the development of spines on medium spiny neurons or the modulation of synaptic plasticity by neurotransmitters.

Applications of human 3D brain cultures

Three-dimensional neural cultures derived from human and other primate pluripotent stem cells are now being used to answer questions about brain development and evolutionary innovation, and to gain insights into human disease.

Human brain development and evolution

The development of the CNS in humans takes a very long time: the generation of astrocytes continues into the first year of life, interneurons migrate for up to two years after birth⁸⁴, and myelination is completed only in the second or third decade of life⁷². Therefore, it is not surprising that knowledge regarding the biology of the stem cell niche, lineage specification and the mechanisms of maturation in human and non-human primate nervous systems is limited. One of the advantages of 3D cultures is that they allow long-term culture (months to years), which could open a window into at least the early stages of human CNS maturation. Forebrain organoids, for which comparison to available data sets from primary brain samples is possible, reach mid-fetal stages of cortical maturation after 3 months *in vitro*^{48,50}. When we maintained such organoids for 20 months *in vitro*, they matured to postnatal stages, as shown by comparison at the single-cell level to primary cortical tissue⁴⁹. More specifically, after about 9–10 months, glial cells switched from an early proliferative state to a mature astrocyte state with different morphologies and physiological effects on neurons⁴⁹. This suggests that there may be an intrinsic molecular

clock that keeps maturation on track, consistent with studies showing that when transplanted into rat or mouse cortex, human cells still take months to mature²².

Compared to other primates, the human cerebral cortex displays a striking expansion, a larger diversity of cortical progenitors, more upper layer neurons and potential differences in cortical interneurons^{72,85}. But why does the development of the human brain take so long, and how do these unique aspects arise? Initial studies with primate-derived 3D cultures suggest that differences in corticogenesis among species may result from cell-autonomous differences in the proliferation of neural progenitors⁸⁶—more specifically, from differences in the cell cycle between humans and chimpanzees⁸⁷. In parallel, studies with neural-crest-derived cells are starting to uncover the regulatory mechanisms that underlie facial development⁸⁸. Comparison of corticogenesis across species in 3D brain cultures will require novel, inventive tools and analytical approaches to capture cell diversity and maturation while accounting for species-related differences in gestation. Assembloids may be particularly relevant in this regard for understanding differences in connectivity, such as the reorganization of corticofugal neurons in primates⁸⁹. Sensory input is absent in these cultures and their cytoarchitecture is still primitive, but it is possible that an organoid system could allow us to study neurons present only in species with larger brains, such as von Economo neurons⁹⁰, or recently identified human-specific subtypes of parvalbumin neurons⁹¹. Genome engineering might even allow us to use organoids to study the effects of genetic variants found in Neanderthal or Denisovan genomes on corticogenesis.

Disease modelling

Human 3D brain cultures already show great promise in modelling monogenic, polygenic and infectious human disorders. Organoids derived from patients with microcephaly who have mutations in the cell cycle-related gene *CDK5RAP2* display an abnormal plane of division in cells located in identifiable ventricular-like zones⁵¹. Organoids in which the tumour suppressor gene *PTEN* is deleted show increased proliferation and delayed neural differentiation, and this phenotype can be manipulated pharmacologically⁵⁷. The 17p13.3 deletion leading to Miller–Dieker syndrome, a severe form of lissencephaly, has been challenging to study

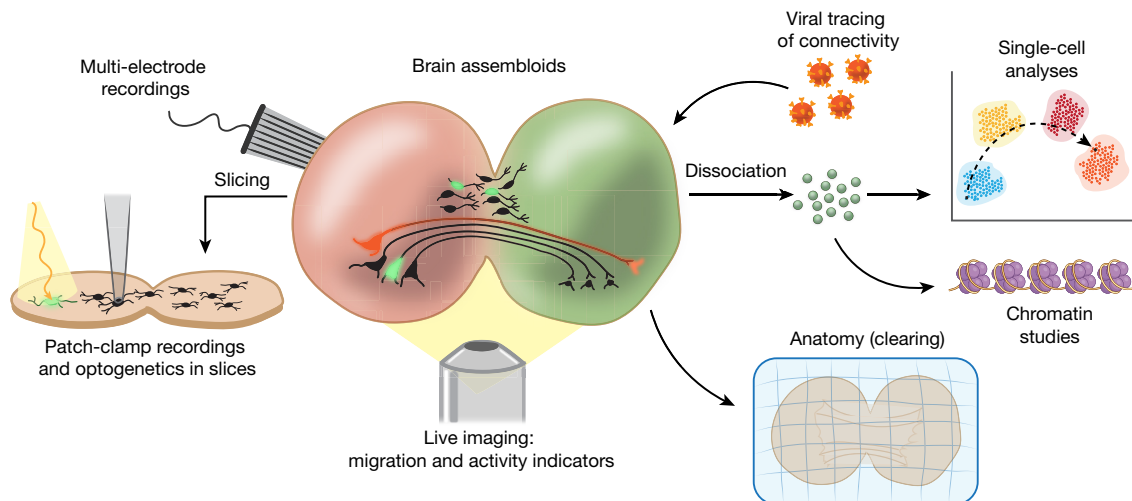


Figure 4 | Methods for probing 3D brain assembloids. Single-cell analyses (for example, transcriptomics, proteomics), chromatin studies (for example, chromatin immunoprecipitation and sequencing (ChIP-seq), assay for transposase-accessible chromatin using sequencing (ATAC-seq)), 3D reconstructions after application of tissue

transparency methods, viral tracing to assess connectivity (retrograde labelling of neurons, red), live imaging of migration and neuronal activity (for example, genetically encoded calcium or voltage indicators), electrophysiology (patch-clamps, multi-electrode recordings) and optogenetic probing in slices or in intact 3D brain assembloids.

in animal models because mice are naturally lissencephalic. Two groups have developed forebrain organoid models of Miller–Dieker syndrome and found abnormalities in radial glia that can be rescued genetically and pharmacologically^{59,60}.

Using human iPS cells from multiple patients with ASD associated with macrocephaly, Mariani *et al.* derived organoids containing both dorsal and ventral domains and found changes in the proportion of GABAergic interneurons⁹². These cortical defects were related to overexpression of the forebrain transcription factor *FOXG1*. In forebrain assembloids derived from patients with Timothy syndrome, a monogenic disease associated with ASD and epilepsy, we found defects in the migration of cortical interneurons that could be restored pharmacologically by modulating the mutated L-type calcium channel⁶⁸. Neurodegenerative disorders have been more challenging to model *in vitro* owing to their late onset, but an early study of 3D human neural cultures carrying Alzheimer’s disease-related mutations showed that they recapitulated both β -amyloid and Tau pathology⁹³.

Gut organoids have already been successfully used to study host–microbe interactions, such as *Helicobacter pylori* infection in gastric organoids⁹⁴. Similarly, human neurons have been essential for identifying the CNS cells that are affected by the microcephaly-related Zika virus infection and for identifying drugs that could reduce infection^{50,95,96}. Cortical organoids derived from patients with a *TREX1*-dependent autoimmune disorder showed increased apoptosis and reduced size, and these abnormalities were mediated by astrocyte-dependent neurotoxicity⁹⁷.

More complex phenotypes in patient-derived cells can be assessed after transplantation into rodents. Hepatocytes, endothelial and stromal cells derived in 3D cultures and subsequently transplanted into mice show expansion of grafts by about 50-fold and secretion of human proteins⁹⁸. Intact 3D brain cultures transplanted into the rodent CNS may integrate better than single-cell suspensions into neural circuits. More importantly, this integration could be shaped by vascularization, interactions with microglia or perhaps even sensory-like activity, and thus offer a unique system for asking questions about higher-level circuit function and dysfunction in disease.

Hurdles and future steps

Brain 3D organoids and assembloids are promising new tools in our arsenal for asking biological and disease-related questions. But there are issues that will need to be addressed. First, brain 3D cultures only approximate the appearance and architecture of neural tissue. They are

smaller (maximum dimensions of 4 mm) than the regions they model, and the internal cytoarchitecture is not always predictable. Radial glia are not attached to a superficial pial surface and although deep and superficial layers separate, it is difficult to derive pristine lamination *in vitro*. White matter regions, which are greatly expanded in primates and contain about two billion neurons in the human brain⁸⁵, are not visible in 3D cultures. Second, specific cell types are either absent, or present in ratios that differ from those found *in vivo*. Microglia, which are of non-ectodermal origin (born in the yolk sack), must migrate into the CNS before the blood–brain barrier closes. Oligodendrocytes are also less abundant in 3D cultures than *in vivo*⁶⁸, but they can be added to 3D ensembles. Similarly, organized meninges and capillaries are not present. Nonetheless, in the absence of circulation, exogenously added endothelial cells may secrete growth factors similar to those secreted in organotypic cultures⁹⁹. Moreover, *in vitro* metabolic demands differ from those *in vivo*, and metabolic needs in the developing brain are species-specific¹⁰⁰. The human brain develops in a low oxygen environment, but we do not know how changes in oxygen tension affect the development of millimetre-wide 3D cultures. Corticogenesis also involves apoptosis. This cell death can release neurotrophic factors¹⁰¹, but it is unclear how cell debris is cleared. Third, human 3D brain cultures lack physiological sensory input and other aspects of developmental plasticity, such as critical periods. What happens, for instance, to corticospinal neurons in long-term cultures in the absence of spinal cord neurons? Future studies will need to address these questions.

Quality control

Predictability is one of the main requirements for disease modelling and drug screening *in vitro* with any platform. Because of a lack of developmental axes and, for some approaches the stochastic differentiation individual 3D brain culture methods should be tested for reproducibility, accuracy, and scalability. It will be important to measure differentiation noise and, as with all dynamic systems, to identify the initial conditions that can drive large effects. Many methods are based on key steps involving Matrigel, which has unpredictable biochemical effects, or the use of up to 10% fetal bovine serum, which varies by supplier or lot and can activate glial cells. How do these conditions affect reproducibility and the reactive state of the cells? How do the cells in 3D brain cultures compare to *in vivo* cell types?¹⁰² Direct comparisons with the developing brain at early stages are relatively straightforward, and most 3D brain cultures map to the first trimester forebrain. But at later stages, when

the proportions of cell types may diverge, these comparisons are challenging and require the comparison of single cells from culture with primary cells⁴⁹. Finally, how scalable and easy to probe are 3D brain cultures? There are certain challenges in obtaining fast readouts in non-homogenous tissue and, for drug screening, penetrability must be considered. Nonetheless, there is evidence that large-scale production and long-term differentiation of 3D brain cultures is possible even in an academic setting^{49,62}.

One approach for addressing these challenges is to introduce quality control steps and to use directed differentiation approaches, which are more predictable. For instance, Arora *et al.*¹⁰³ used an automated micropipette system to identify intestinal pre-organoids from hindgut cultures. Size and morphology, expression of selected markers, live dyes for survival or brain-region-specific reporters can each be used in different combinations as read-outs. In disease models, a large fraction of the variance is driven by inter-individual differences^{104–106} and therefore large sample sizes and the use of isogenic hPS cell lines is essential.

Tools and biomaterials

Another strategy for increasing predictability, recapitulating key developmental features *in vitro* and obtaining reliable read-outs is to develop new biomaterial approaches and to apply novel tools for probing and manipulating brain organoids or assembloids (Fig. 4).

Light-sheet microscopy, which is fast and causes minimal photo-damage, as well as other advanced microscopy techniques combined with genetically encoded calcium or voltage indicators, have the potential to capture the activity of whole 3D brain cultures over long periods of time. These imaging technologies can also be used to investigate tissue self-organization, activity waves during early development and the early emergence of disease phenotypes^{107,108}. Highly parallel single-cell transcriptomics, and soon large-scale proteomics, will be able to provide insights into the diversity of cell types and cell states, lineage progression and defects in differentiation in patients and across species^{49,53,54,87}. Tissue transparency methods, such as CLARITY¹⁰⁹, as well as anterograde and retrograde viral labelling techniques¹¹⁰, can be used to map neuronal connectivity. Last, electrophysiological recordings with multi-electrode recordings or in slices combined with optogenetic techniques can capture network dynamics, including the potential emergence of neural oscillations. Acquiring multi-level read-outs on a large scale will require the use of advanced mathematical tools to comprehend self-organization principles and the cytodynamics underlying these complex processes, and to reliably identify disease-related phenotypes.

Novel biomaterials are required for both controlling neural patterning and for breaking symmetry in a predictable way, but also to advance the maturation and scalability of these cultures and the emergence of inaccessible biological processes. Stiffness influences morphogenesis and differentiation. Bio-scaffolds can also compartmentalize space, as shown for 3D cultures of salivary gland¹¹¹ or lacrimal gland¹¹². The ECM has a unique composition in the developing human brain¹¹³ and the perineuronal nets are thought to regulate neural plasticity¹¹⁴. Most of the biomaterials used to date for organoid cultures are insufficiently defined and have unpredictable effects on differentiation. In this regard, hydrogels, which are hydrophilic polymers that can be generated using a large variety of natural or synthetic materials (for example, poly-ethylene-glycol (PEG) or poly-vinyl-alcohol) hold great promise. Geometric confinement of hPS cells to PEG-patterned substrates, for instance, facilitates self-organization of cardiac lineages and results in beating cardiac microchambers¹¹⁵. Hydrogels are programmable, which is important for brain cultures where reproducing the non-uniform environment is probably more essential than scaling up. By manipulating their pore size and topology (that is, void space) and physical properties (for example, elasticity and topography), hydrogels can be assembled into higher-order architectures¹¹⁶. Mechanical forces and cell patterning can be modulated locally. 'Writing', soft lithography and other bioprinting strategies¹¹⁷ can be used to achieve 3D micropatterning by embedding hydrogels with particles that release or sequester small molecules, growth factors,

aptamers, nanoparticles or active peptide sequences. This compartmentalization could create transient organizers and morphogen gradients. Next-generation hydrogels and synthetic ECM will need to improve cell viability over larger scales (up to centimetres), incorporate dynamic features such as pH and oxygen sensing, and eliminate toxic agents.

We still know little about ECM in the developing human brain, but a reverse engineering approach could be used to de-cellularize brain tissue and use hydrogels to derive physiological scaffolds for 3D neural differentiation. Subsequently, ECM components that are necessary for deriving specific features in 3D brain cultures can be used to generate synthetic biomaterials and increase scalability. These experiments could also be informative for achieving predictable self-organization of specific brain regions.

Novel features in human 3D brain cultures

The combination of biomaterials and state-of-the-art technologies for manipulating human 3D brain cultures has the potential to give rise to novel features *in vitro* and accelerate the study of human brain development and disease. A more permissive environment and extensive growth may lead to a deeper understanding of cortical folding and clarify how size is coupled with timing. For instance, will maturation depend on ensuring larger sizes of tissue *in vitro*, providing external stimulation, or achieving myelination? This is important because accelerating functional maturation up to later stages of postnatal development in brain assembloids could facilitate the study of neural circuits and help us understand how neural oscillations arise, and could ultimately inform models of neuro-degeneration. As these and other features emerge and more elaborate transplantations in rodents and other species are being planned, discussions on the ethical aspects of this work should be pursued^{118–120}. Engaging the public using accurate descriptions, for example by avoiding the use of terms such as 'mini-brains', will be essential.

Outlook

This is an exciting new field and as with many technologies, it may follow a 'hype' cycle¹²¹ in which we overestimate its effects in the short run and underestimate its effects in the long run. A better understanding of the complexity of this platform, and bringing interdisciplinary approaches will accelerate our progress up a 'slope of enlightenment' and into the 'plateau of productivity'.

Received 6 September; accepted 14 November 2017.

- Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
- Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Zhang, S. C., Wernig, M., Duncan, I. D., Brustle, O. & Thomson, J. A. *In vitro* differentiation of transplantable neural precursors from human embryonic stem cells. *Nat. Biotechnol.* **19**, 1129–1133 (2001).
- Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
- Zhang, Y. *et al.* Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
- Heidenreich, M. & Zhang, F. Applications of CRISPR–Cas systems in neuroscience. *Nat. Rev. Neurosci.* **17**, 36–44 (2016).
- Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
- Box, G. E. P. Science and statistics. *J. Am. Stat. Assoc.* **71**, 791–799 (1976).
- Kant, I. *Critique of Judgment* (1790).
- Townes, P. L. & Holtfrete, J. Directed movements and selective adhesion of embryonic amphibian cells. *J. Exp. Zool.* **128**, 53–120 (1955).
- Gierer, A. *et al.* Regeneration of hydra from reaggregated cells. *Nat. New Biol.* **239**, 98–101 (1972).
- Tsaiaris, C. D. & Aulehla, A. Self-organization of embryonic genetic oscillators into spatiotemporal wave patterns. *Cell* **164**, 656–667 (2016).
- Sun, T. & Hevner, R. F. Growth and folding of the mammalian cerebral cortex: from molecules to malformations. *Nat. Rev. Neurosci.* **15**, 217–232 (2014).
- Paşca, S. P., Panagiotakos, G. & Dolmetsch, R. E. Generating human neurons *in vitro* and using them to understand neuropsychiatric disease. *Annu. Rev. Neurosci.* **37**, 479–501 (2014).

16. Warmflash, A., Sorre, B., Etoc, F., Siggia, E. D. & Brivanlou, A. H. A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. *Nat. Methods* **11**, 847–854 (2014).
17. Turing, A. M. The chemical basis of morphogenesis. *Philos. Trans. R. Soc. B* **237**, 37–72 (1952).
18. Deglincerti, A., Etoc, F., Ozair, M. Z. & Brivanlou, A. H. Self-organization of spatial patterning in human embryonic stem cells. *Curr. Top. Dev. Biol.* **116**, 99–113 (2016).
19. Holtfrete, J. Neural differentiation of ectoderm through exposure to saline solution. *J. Exp. Zool.* **95**, 307–344 (1944).
20. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280 (2009).
- This paper introduces a small-molecule-based approach for neural induction of hPS cells.**
21. Elkabetz, Y. *et al.* Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes Dev.* **22**, 152–165 (2008).
22. Espuny-Camacho, I. *et al.* Pyramidal neurons derived from human pluripotent stem cells integrate efficiently into mouse brain circuits *in vivo*. *Neuron* **77**, 440–456 (2013).
23. Gaspard, N. *et al.* An intrinsic mechanism of corticogenesis from embryonic stem cells. *Nature* **455**, 351–357 (2008).
24. Paşca, S. P. *et al.* Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. *Nat. Med.* **17**, 1657–1662 (2011).
25. Pankratz, M. T. *et al.* Directed neural differentiation of human embryonic stem cells via an obligated primitive anterior stage. *Stem Cells* **25**, 1511–1520 (2007).
26. Kriks, S. *et al.* Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. *Nature* **480**, 547–551 (2011).
27. Reddington, A. E., Rosser, A. E. & Dunnett, S. B. Differentiation of pluripotent stem cells into striatal projection neurons: a pure MSN fate may not be sufficient. *Front. Cell. Neurosci.* **8**, 398 (2014).
28. Hirano, M. *et al.* Generation of structures formed by lens and retinal cells differentiating from embryonic stem cells. *Dev. Dyn.* **228**, 664–671 (2003).
29. Kawaguchi, K., Kageyama, R. & Sano, M. Topological defects control collective dynamics in neural progenitor cell cultures. *Nature* **545**, 327–331 (2017).
30. Pickl, M. & Ries, C. H. Comparison of 3D and 2D tumor models reveals enhanced HER2 activation in 3D associated with an increased response to trastuzumab. *Oncogene* **28**, 461–468 (2009).
31. Turner, D. A., Baillie-Johnson, P. & Martinez Arias, A. Organoids and the genetically encoded self-assembly of embryonic stem cells. *BioEssays* **38**, 181–191 (2016).
32. Nakano, T. *et al.* Self-formation of optic cups and storable stratified neural retina from human ESCs. *Cell Stem Cell* **10**, 771–785 (2012).
33. Eiraku, M. *et al.* Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* **472**, 51–56 (2011).
- Work in refs 32 and 33 recapitulates the morphogenesis of the optic cup in 3D cultures generated from hPS cells or mPS cells.**
34. Takasato, M. *et al.* Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis. *Nature* **526**, 564–568 (2015).
35. Chen, Y. W. *et al.* A three-dimensional model of human lung development and disease from pluripotent stem cells. *Nat. Cell Biol.* **19**, 542–549 (2017).
36. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
- This paper demonstrates that single intestinal stem cells suspended in an extracellular matrix can generate crypt-villus organoids *in vitro*.**
37. Harrison, R. G. Observations on the living developing nerve fiber. *Proc. Soc. Exp. Biol. Med.* **4**, 140–143 (1907).
38. Humpel, C. Organotypic brain slice cultures: A review. *Neuroscience* **305**, 86–98 (2015).
39. Henschen, A., Hoffer, B. & Olson, L. Spinal cord grafts in oculo: survival, growth, histological organization and electrophysiological characteristics. *Exp. Brain Res.* **60**, 38–47 (1985).
40. Reynolds, B. A., Tetzlaff, W. & Weiss, S. A multipotent EGF-responsive striatal embryonic progenitor cell produces neurons and astrocytes. *J. Neurosci.* **12**, 4565–4574 (1992).
41. Phan, D. T. *et al.* Blood-brain barrier-on-a-chip: Microphysiological systems that capture the complexity of the blood-central nervous system interface. *Exp. Biol. Med.* **242**, 1669–1678 (2017).
42. Sachs, N., Tsukamoto, Y., Kujala, P., Peters, P. J. & Clevers, H. Intestinal epithelial organoids fuse to form self-organizing tubes in floating collagen gels. *Development* **144**, 1107–1112 (2017).
43. Sato, T. *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762–1772 (2011).
44. Huch, M. *et al.* Unlimited *in vitro* expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis. *EMBO J.* **32**, 2708–2721 (2013).
45. Watanabe, K. *et al.* Directed differentiation of telencephalic precursors from embryonic stem cells. *Nat. Neurosci.* **8**, 288–296 (2005).
46. Eiraku, M. *et al.* Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. *Cell Stem Cell* **3**, 519–532 (2008).
47. Mariani, J. *et al.* Modeling human cortical development *in vitro* using induced pluripotent stem cells. *Proc. Natl Acad. Sci. USA* **109**, 12770–12775 (2012).
48. Paşca, A. M. *et al.* Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nat. Methods* **12**, 671–678 (2015).
49. Sloan, S. A. *et al.* Human astrocyte maturation captured in 3D cerebral cortical spheroids derived from pluripotent stem cells. *Neuron* **95**, 779–790 (2017).
50. Qian, X. *et al.* Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell* **165**, 1238–1254 (2016).
- Refs 46–50 and 58 describe the development of brain-region-specific 3D cultures.**
51. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
- This paper develops organoids that resemble various brain regions.**
52. Renner, M. *et al.* Self-organized developmental patterning and differentiation in cerebral organoids. *EMBO J.* **36**, 1316–1329 (2017).
53. Camp, J. G. *et al.* Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl Acad. Sci. USA* **112**, 15672–15677 (2015).
54. Quadrato, G. *et al.* Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* **545**, 48–53 (2017).
55. Lancaster, M. A. *et al.* Guided self-organization and cortical plate formation in human brain organoids. *Nat. Biotechnol.* **35**, 659–666 (2017).
56. Bagley, J. A., Reumann, D., Bian, S., Lévi-Strauss, J. & Knoblich, J. A. Fused cerebral organoids model interactions between brain regions. *Nat. Methods* **14**, 743–751 (2017).
57. Li, Y. *et al.* Induction of expansion and folding in human cerebral organoids. *Cell Stem Cell* **20**, 385–396 (2017).
58. Kadoshima, T. *et al.* Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc. Natl Acad. Sci. USA* **110**, 20284–20289 (2013).
59. Bershteyn, M. *et al.* Human iPSC-derived cerebral organoids model cellular features of lissencephaly and reveal prolonged mitosis of outer radial glia. *Cell Stem Cell* **20**, 435–449 (2017).
60. Iefremova, V. *et al.* An organoid-based model of cortical development identifies non-cell-autonomous defects in Wnt signaling contributing to Miller-Dieker syndrome. *Cell Reports* **19**, 50–59 (2017).
61. Sakaguchi, H. *et al.* Generation of functional hippocampal neurons from self-organizing human embryonic stem cell-derived dorsomedial telencephalic tissue. *Nat. Commun.* **6**, 8896 (2015).
62. Rigamonti, A. *et al.* Large-scale production of mature neurons from human pluripotent stem cells in a three-dimensional suspension culture system. *Stem Cell Reports* **6**, 993–1008 (2016).
63. Jo, J. *et al.* Midbrain-like organoids from human pluripotent stem cells contain functional dopaminergic and neuromelanin-producing neurons. *Cell Stem Cell* **19**, 248–257 (2016).
64. Zecca, L. *et al.* The neuromelanin of human substantia nigra: structure, synthesis and molecular behaviour. *J. Neural Transm. Suppl.* **65**, 145–155 (2003).
65. Marton, R. M. & Paşca, S. P. Neural differentiation in the third dimension: generating a human midbrain. *Cell Stem Cell* **19**, 145–146 (2016).
66. Suga, H. *et al.* Self-formation of functional adenohipophysis in three-dimensional culture. *Nature* **480**, 57–62 (2011).
67. Muguruma, K., Nishiyama, A., Kawakami, H., Hashimoto, K. & Sasai, Y. Self-organization of polarized cerebellar tissue in 3D culture of human pluripotent stem cells. *Cell Reports* **10**, 537–550 (2015).
68. Birey, F. *et al.* Assembly of functionally integrated human forebrain spheroids. *Nature* **545**, 54–59 (2017).
- This work develops human forebrain assembloids to study interneuron migration and circuit formation and to model disease.**
69. Kepecs, A. & Fishell, G. Interneuron cell types are fit to function. *Nature* **505**, 318–326 (2014).
70. Bartolini, G., Cicci, G. & Marín, O. Integration of GABAergic interneurons into cortical cell assemblies: lessons from embryos and adults. *Neuron* **79**, 849–864 (2013).
71. Wonders, C. P. & Anderson, S. A. The origin and specification of cortical interneurons. *Nat. Rev. Neurosci.* **7**, 687–696 (2006).
72. Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The cellular and molecular landscapes of the developing human central nervous system. *Neuron* **89**, 248–268 (2016).
73. Anderson, S. A., Eisenstat, D. D., Shi, L. & Rubenstein, J. L. Interneuron migration from basal forebrain to neocortex: dependence on *Dlx* genes. *Science* **278**, 474–476 (1997).
74. Rubenstein, J. L. & Merzenich, M. M. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* **2**, 255–267 (2003).
75. Marín, O. Interneuron dysfunction in psychiatric disorders. *Nat. Rev. Neurosci.* **13**, 107–120 (2012).
76. Xiang, Y. *et al.* Fusion of regionally specified hPSC-derived organoids models human brain development and interneuron migration. *Cell Stem Cell* **21**, 383–398 (2017).
77. Schwartz, M. P. *et al.* Human pluripotent stem cell-derived neural constructs for predicting neural toxicity. *Proc. Natl Acad. Sci. USA* **112**, 12516–12521 (2015).
78. Sternfeld, M. J. *et al.* Speed and segmentation control mechanisms characterized in rhythmically-active circuits created from spinal neurons produced from genetically-tagged embryonic stem cells. *eLife* **6**, e21540 (2017).
79. Workman, M. J. *et al.* Engineered human pluripotent-stem-cell-derived intestinal tissues with a functional enteric nervous system. *Nat. Med.* **23**, 49–59 (2017).

This paper describes the assembly of intestinal and neural tissue in 3D cultures.

80. Yamamoto, N., Yamada, K., Kurotani, T. & Toyama, K. Laminar specificity of extrinsic cortical connections studied in coculture preparations. *Neuron* **9**, 217–228 (1992).
81. Bolz, J., Novak, N. & Staiger, V. Formation of specific afferent connections in organotypic slice cultures from rat visual cortex cocultured with lateral geniculate nucleus. *J. Neurosci.* **12**, 3054–3070 (1992).
82. Woolsey, T. A. & Van der Loos, H. The structural organization of layer IV in the somatosensory region (SI) of mouse cerebral cortex. The description of a cortical field composed of discrete cytoarchitectonic units. *Brain Res.* **17**, 205–242 (1970).
83. Chun, S. *et al.* Specific disruption of thalamic inputs to the auditory cortex in schizophrenia models. *Science* **344**, 1178–1182 (2014).
84. Paredes, M. F. *et al.* Extensive migration of young neurons into the infant human frontal lobe. *Science* **354**, aaf7073 (2016).
85. Sousa, A. M. M., Meyer, K. A., Santpere, G., Gulden, F. O. & Sestan, N. Evolution of the human nervous system function, structure, and development. *Cell* **170**, 226–247 (2017).
86. Otani, T., Marchetto, M. C., Gage, F. H., Simons, B. D. & Livesey, F. J. 2D and 3D stem cell models of primate cortical development identify species-specific differences in progenitor behavior contributing to brain size. *Cell Stem Cell* **18**, 467–480 (2016).
87. Mora-Bermúdez, F. *et al.* Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. *eLife* **5**, e18683 (2016).
88. Prescott, S. L. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).
89. Heffner, R. S. & Masterton, R. B. The role of the corticospinal tract in the evolution of human digital dexterity. *Brain Behav. Evol.* **23**, 165–183 (1983).
90. Defelipe, J. The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. *Front. Neuroanat.* **5**, 29 (2011).
91. Luo, C. *et al.* Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).
92. Mariani, J. *et al.* FOXG1-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders. *Cell* **162**, 375–390 (2015).
- This work identifies forebrain defects in organoids derived from patients with autism spectrum disorders.**
93. Choi, S. H. *et al.* A three-dimensional human neural cell culture model of Alzheimer's disease. *Nature* **515**, 274–278 (2014).
94. McCracken, K. W. *et al.* Modelling human development and disease in pluripotent stem-cell-derived gastric organoids. *Nature* **516**, 400–404 (2014).
95. Gabriel, E. *et al.* Recent Zika virus isolates induce premature differentiation of neural progenitors in human brain organoids. *Cell Stem Cell* **20**, 397–406 (2017).
96. Cugola, F. R. *et al.* The Brazilian Zika virus strain causes birth defects in experimental models. *Nature* **534**, 267–271 (2016).
97. Thomas, C. A. *et al.* Modeling of TREX1-dependent autoimmune disease using human stem cells highlights L1 accumulation as a source of neuroinflammation. *Cell Stem Cell* **21**, 319–331 (2017).
98. Stevens, K. R. *et al.* *In situ* expansion of engineered human liver tissue in a mouse model of chronic liver disease. *Sci. Transl. Med.* **9**, eaah5505 (2017).
99. Hutter-Schmid, B., Kniewallner, K. M. & Humpel, C. Organotypic brain slice cultures as a model to study angiogenesis of brain vessels. *Front. Cell Dev. Biol.* **3**, 52 (2015).
100. Erecinska, M., Cherian, S. & Silver, I. A. Energy metabolism in mammalian brain during development. *Prog. Neurobiol.* **73**, 397–445 (2004).
101. Agasse, F., Roger, M. & Coronas, V. Neurogenic and intact or apoptotic non-neurogenic areas of adult brain release diffusible molecules that differentially modulate the development of subventricular zone cell cultures. *Eur. J. Neurosci.* **19**, 1459–1468 (2004).
102. Stein, J. L. *et al.* A quantitative framework to evaluate modeling of cortical development by neural stem cells. *Neuron* **83**, 69–86 (2014).
103. Arora, N. *et al.* A process engineering approach to increase organoid yield. *Development* **144**, 1128–1136 (2017).
104. Carcamo-Orive, I. *et al.* Analysis of transcriptional variability in a large human iPSC library reveals genetic and non-genetic determinants of heterogeneity. *Cell Stem Cell* **20**, 518–532 (2017).
105. Kilpinen, H. *et al.* Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370–375 (2017).
106. Rouhani, F. *et al.* Genetic background drives transcriptional variation in human induced pluripotent stem cells. *PLoS Genet.* **10**, e1004432 (2014).
107. Lin, M. Z. & Schnitzer, M. J. Genetically encoded indicators of neuronal activity. *Nat. Neurosci.* **19**, 1142–1153 (2016).
108. Kim, S. Y., Chung, K. & Deisseroth, K. Light microscopy mapping of connections in the intact brain. *Trends Cogn. Sci.* **17**, 596–599 (2013).
109. Chung, K. & Deisseroth, K. CLARITY for mapping the nervous system. *Nat. Methods* **10**, 508–513 (2013).
110. Callaway, E. M. & Luo, L. Monosynaptic circuit tracing with glycoprotein-deleted rabies viruses. *J. Neurosci.* **35**, 8979–8985 (2015).
111. Ogawa, M. *et al.* Functional salivary gland regeneration by transplantation of a bioengineered organ germ. *Nat. Commun.* **4**, 2498 (2013).
112. Hirayama, M. *et al.* Functional lacrimal gland regeneration by transplantation of a bioengineered organ germ. *Nat. Commun.* **4**, 2497 (2013).
113. Fietz, S. A. *et al.* Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *Proc. Natl Acad. Sci. USA* **109**, 11836–11841 (2012).
114. Sorg, B. A. *et al.* Casting a wide net: role of perineuronal nets in neural plasticity. *J. Neurosci.* **36**, 11459–11468 (2016).
115. Ma, Z. *et al.* Self-organizing human cardiac microchambers mediated by geometric confinement. *Nat. Commun.* **6**, 7413 (2015).
116. Chiang, M. Y., Hsu, Y. W., Hsieh, H. Y., Chen, S. Y. & Fan, S. K. Constructing 3D heterogeneous hydrogels from electrically manipulated prepolymer droplets and crosslinked microgels. *Sci. Adv.* **2**, e1600964 (2016).
117. Deforest, C. A., Sims, E. A. & Anseth, K. S. Peptide-functionalized click hydrogels with independently tunable mechanics and chemical functionality for 3D cell culture. *Chem. Mater.* **22**, 4783–4790 (2010).
118. Aach, J., Lunshof, J., Iyer, E. & Church, G. M. Addressing the ethical issues raised by synthetic human entities with embryo-like features. *eLife* **6**, e20674 (2017).
119. Munsie, M., Hyun, I. & Sugarman, J. Ethical issues in human organoid and gastruloid research. *Development* **144**, 942–945 (2017).
120. Pera, M. F. *et al.* What if stem cells turn into embryos in a dish? *Nat. Methods* **12**, 917–919 (2015).
121. Mason, C. & Manzotti, E. Induced pluripotent stem cells: an emerging technology platform and the Gartner hype cycle. *Regen. Med.* **4**, 329–331 (2009).

Acknowledgements The author thanks B. A. Barres, K. Deisseroth, R. Reimer, J. A. Bernstein, H. B. Fraser, J. R. Huguenard and members of the Pasca laboratory, and acknowledges funding support from the National Institute of Mental Health (NIMH), California Institute of Regenerative Medicine (CIRM), the MQ Fellow Award, the Donald E. and Delia B. Baxter Foundation Award, the Stanford Neurosciences Institute's Brain Rejuvenation Project, the Kwan Research Fund and the Brain & Behavior Research Foundation (BBRF, NARSAD).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to S.P.P. (spasca@stanford.edu).

The biology and management of non-small cell lung cancer

Roy S. Herbst¹, Daniel Morgensztern² & Chris Boshoff^{1,3}

Important advancements in the treatment of non-small cell lung cancer (NSCLC) have been achieved over the past two decades, increasing our understanding of the disease biology and mechanisms of tumour progression, and advancing early detection and multimodal care. The use of small molecule tyrosine kinase inhibitors and immunotherapy has led to unprecedented survival benefits in selected patients. However, the overall cure and survival rates for NSCLC remain low, particularly in metastatic disease. Therefore, continued research into new drugs and combination therapies is required to expand the clinical benefit to a broader patient population and to improve outcomes in NSCLC.

Lung cancer is the most common cause of cancer death worldwide, with an estimated 1.6 million deaths each year¹. Approximately 85% of patients have a group of histological subtypes collectively known as NSCLC, of which lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the most common subtypes². The most common aetiology for lung cancer is tobacco smoking, accounting for more than 80% of cases in the United States and other countries where smoking is common³. Although all major histological subtypes of NSCLC, as well as small cell lung cancer (SCLC), are associated with smoking, the association is stronger with LUSC and SCLC than LUAD, with the latter being the most common histology in never smokers⁴. Lung cancer in never smokers is more common in women and in East Asia, and has been associated with environmental exposures including second-hand smoking, pollution and occupational carcinogens, and with inherited genetic susceptibility^{3,5,6}.

Eradicating the use of all tobacco-related products is a key goal of the global fight against cancer and requires a comprehensive approach. Primary prevention efforts include targeting nicotine addiction by providing effective delivery of nicotine without the co-administration of carcinogenic chemicals that are present in cigarettes, such as via e-cigarettes⁷. Other strategies include the use of varenicline, a partial agonist of the nicotinic acetylcholine receptor⁸, counselling and other socio-economic methods including taxation, advertisement and legislative measures, such as lowering the amount of nicotine in cigarettes to non-addictive levels, a policy recently announced by the US Food and Drug Administration (FDA) (<https://www.fda.gov/tobaccoproducts/newsevents/ucm568425.htm>). Despite best intentions, the use of e-cigarettes to facilitate smoking cessation remains unproven and controversial given concerns that they might promote the initiation of new individuals to use smoking devices with unknown long-term consequences⁹.

Although a crucial component of the fight against lung cancer, tobacco prevention strategies are not enough to win the war. Increasingly sophisticated therapies are required to meaningfully improve clinical outcomes for patients. Progress in this area has been substantial and promising over the past 20 years with the advent of various targeted therapies and the effective application of immunotherapy in some populations of patients with advanced NSCLC. Yet, major challenges still remain, including the identification of new driver gene alterations to expand the population that benefit from targeted therapies, better understanding of mechanisms of resistance to targeted therapy to allow them to be prevented or overcome,

and the need for better predictors of responses to immunotherapy, new drugs and rationally designed drug combination therapies. In this Review, we provide an overview of the recent progress made in lung cancer biology and treatment, including the most promising strategies that have already made a notable impact in outcomes for patients with advanced stage NSCLC.

Biology of lung cancer

Lung cancer is a molecularly heterogeneous disease and understanding its biology is crucial for the development of effective therapies. The treatment of lung cancer has changed from the empirical use of cytotoxic therapy based on a physician's preference to a hallmark of personalized medicine, with subsets of patients treated according to the genetic alterations of their tumour and the status of programmed death ligand-1 (PD-L1), which predict for benefit from targeted therapies or immune checkpoint blockers (ICBs), respectively.

Similar to most malignancies, lung cancer is composed of sub-populations of cells, or clones, with distinct molecular features, resulting in intra-tumoral heterogeneity. A larger sub-clonal mutation fraction may be associated with increased likelihood of postsurgical relapse in patients with localized LUAD, implying that there is a greater propensity for metastases early during tumour development in those tumours with increased intra-tumoral heterogeneity¹⁰. The identification of clonal targetable genetic alterations occurring early during cancer evolution has changed the paradigm of treatment for oncogene-addicted cancers, although few patients, if any, are cured owing to inherent and acquired resistance mechanisms, with the latter occurring mostly through the selection of resistant sub-clones, pre-existing before targeted drug exposure^{11,12}.

The most common genetic alterations in LUAD and LUSC are shown in Box 1. Variant allele frequencies for somatic mutations have shown that mutations in the Kirsten rat sarcoma (*KRAS*) and epidermal growth factor receptor (*EGFR*) genes, when detected, are usually present in the founder clones, indicating their roles in tumour initiation and representing attractive targets for therapeutic intervention. *KRAS* and *EGFR* mutations are usually mutually exclusive, but when they co-exist, *KRAS* mutations may confer resistance to *EGFR* inhibitors¹³. Tumours that harbour oncogenic drivers such as *EGFR* mutations¹⁴ and *ROS1* and anaplastic lymphoma kinase (*ALK*) rearrangements have a lower-than-average mutation load, mostly owing to occurrence in never or light smokers, although the dominant driving properties of these oncogenes may reduce the selection pressure for acquiring additional mutations.

¹Yale Cancer Center, Yale School of Medicine, New Haven, Connecticut, USA. ²Washington University School of Medicine, St Louis, Missouri, USA. ³Pfizer, Inc. New York City, New York, USA.

BOX 1

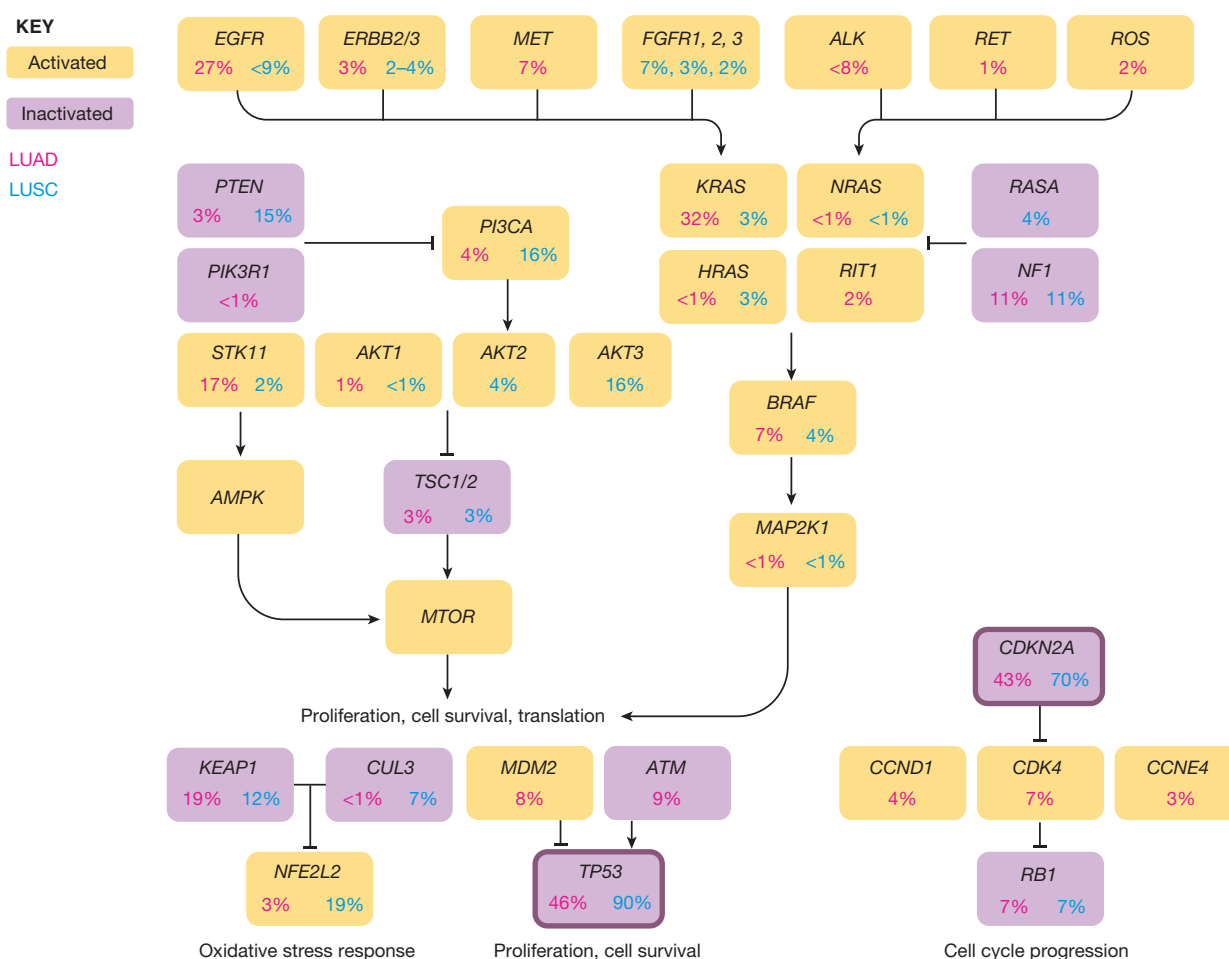
Alterations in targetable oncogenic pathways in LUAD and LUSC

Pathway diagram showing the percentage of NSCLC with alterations involving key pathway components for receptor tyrosine kinase signalling, mTOR signalling, oxidative stress response, proliferation and cell cycle progression. The frequency of alterations is based on the sum of somatic mutations, homozygous deletions, focal amplifications, and by significant up- or downregulation of gene expression (for example, *AKT3*, *FGFR1*, *PTEN*).

The most commonly mutated genes in LUAD include *KRAS* and *EGFR*, and the tumour suppressor genes *TP53*, *KEAP1*, *STK11* and *NF1*. The frequency of *EGFR*-activating mutations varies greatly by region and ethnicity. *KEAP1* inactivation in the presence of *KRAS* mutations confers sensitivity to inhibition of glutaminase in

preclinical lung cancer models, providing a potential therapeutic strategy in dual *KEAP1*- and *KRAS*-mutant LUAD¹³⁹.

Common mutated genes in LUSC include the tumour suppressors *TP53*, which is present in more than 90% of tumours, and *CDKN2A*. The latter, which encodes the p16^{INK4A} and p14^{ARF} proteins, is inactivated in over 70% of LUSC through epigenetic silencing by methylation (21%), inactivating mutation (18%), exon 1 β skipping (4%), or homozygous deletion (29%). Although *EGFR* amplification occurs, unlike LUAD, actionable mutations in receptor tyrosine kinases are rarely observed in LUSC. (Data compiled from refs 14, 17, 22, 45 and diagram adapted from refs 17, 22.)



Unlike LUAD, actionable mutations in receptor tyrosine kinases are rarely detected in LUSC¹⁵.

Mutations in tumour protein p53 (*TP53*) are more commonly observed with advancing grade, suggesting a role during tumour progression¹⁶. By contrast, the frequency of *KRAS* mutations in LUAD seems constant across tumour grades, suggesting a role in tumour initiation or early tumorigenesis, and supporting the presence of *KRAS* alterations in founder clones.

The genomic landscape of lung cancer is markedly distinct between never smokers and smokers, with the latter containing a significantly higher mutation frequency, predominantly cytosine to adenine (C>A) nucleotide transversions and non-actionable mutations such as those in

KRAS and *TP53*. By contrast, never smokers usually have a predominant transition of cytosine to thymine (C>T), and a higher prevalence of actionable driving gene alterations including activating *EGFR* mutations, and *ROS1* and *ALK* translocations^{14,17}.

Tumour microenvironment

Genetic events that initiate and drive tumour evolution also shape the tumour microenvironment (TME). Therefore, the genetic architecture of a tumour determines not only the fitness of the cancer cells, but also the composition of the TME. NSCLC has a particularly high somatic tumour mutation burden (TMB), defined as the number of nonsynonymous coding mutations per megabase, particularly in smokers, who represent

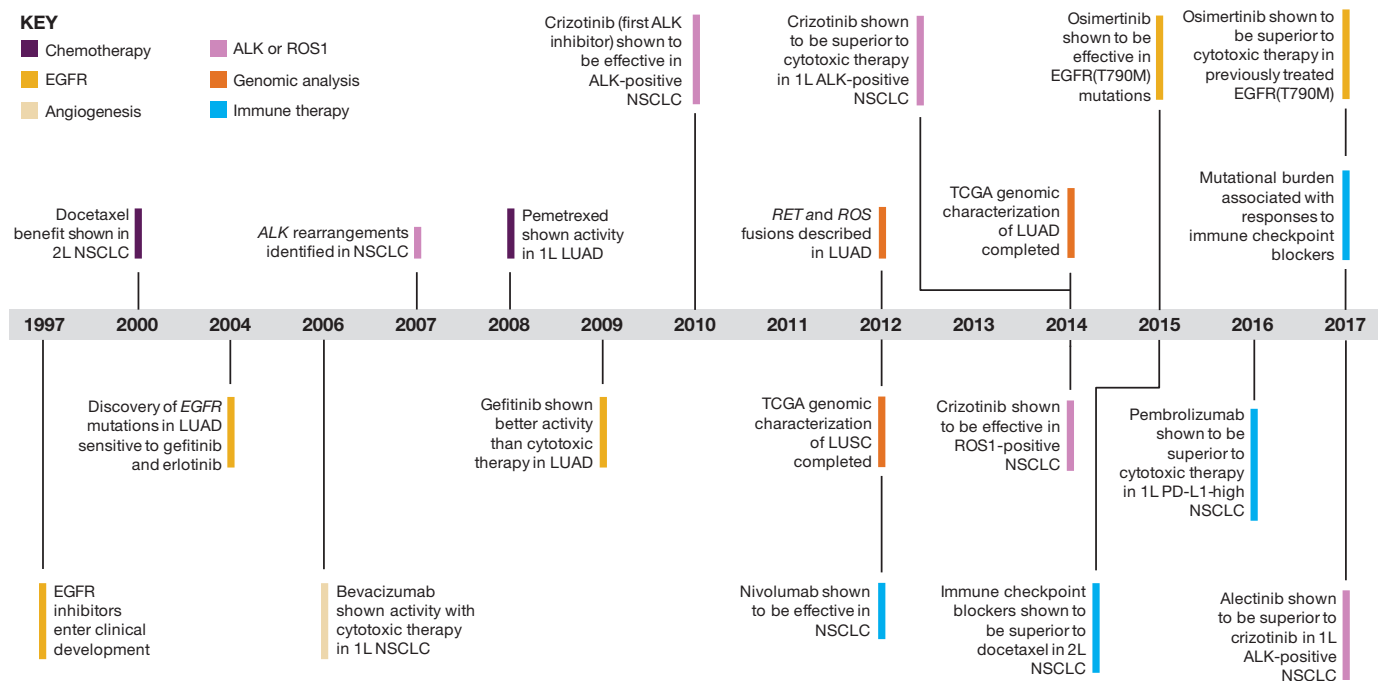


Figure 1 | Timeline illustrating the development of targeted therapies and immunotherapies for the treatment of NSCLC over two decades. Timeline highlights some of the pivotal discoveries and clinical studies that

have transformed the management of NSCLC over the past two decades. 1L, first line; 2L, second line; TCGA, The Cancer Genome Atlas.

the majority of patients. Overall, the number of mutations is significantly higher in metastases than in primary lung lesions¹⁸.

Some mutations create neoantigens, which may be recognized by tumour-infiltrating cytotoxic T cells. A high clonal neoantigen burden in LUAD is associated with an inflamed TME, enriched with activated effector T cells and the expression of proteins associated with antigen presentation, T cell migration (CXCL-10 and CXCL-9), and effector T cell function, as well as negative regulators of T-cell activity including PD-L1, programmed death-1 (PD-1) and lymphocyte activation gene-3 (LAG-3)¹⁹. This phenotype may confer sensitivity to treatment with an ICB. Loss of mismatch-repair function, which confers the microsatellite instability phenotype, is an extreme example of cancer with a high TMB, with these tumours demonstrating T cell infiltration and marked responses to ICBs, independent of the tissue of origin²⁰.

Genetic alterations may affect the TME in several ways. An example is the inactivation of the tumour suppressor serine/threonine kinase 11 (*STK11*; also known as *LKB1*), occurring in one-third of *KRAS*-mutated LUAD, which skews the TME towards the accumulation of immunosuppressive neutrophils and loss of PD-L1 expression, and is associated with fewer tumour-infiltrating lymphocytes²¹. Large-scale studies to correlate the genomes of NSCLC to the cellular constituents of the TME are required to understand how different genotypes determine the cellular make-up of the TME.

Therapeutic advances during the past two decades

Over the past 20 years, treatment has evolved from the empiric use of cytotoxic therapies to effective and better tolerated regimens that are targeted to specific molecular subtypes in LUAD, and therapies in development to target LUSC^{17,22} (Fig. 1). Platinum-based doublet therapy (for example, cisplatin in combination with another cytotoxic therapy) has been the standard therapy for patients with advanced stage NSCLC and good performance status, with the option of maintenance therapy in patients with non-LUSC histology who achieve tumour control after the initial four to six cycles²³. Overall, there have been no clinically meaningful differences in outcome among the multiple cytotoxic regimens used in patients with advanced stage NSCLC²⁴, with the exception of pemetrexed, which is less effective in patients with LUSC²⁵. The addition

of bevacizumab, a monoclonal antibody against vascular endothelial growth factor (VEGF) or necitumumab, an antibody that targets EGFR, produced modest improvements in survival for patients with non-LUSC and LUSC histological subtypes, respectively^{26,27}. In the second-line setting, the standard of care has been docetaxel, with or without the anti-VEGF receptor-2 antibody ramucirumab²⁸.

Cytotoxic regimens have demonstrated their greatest effect on earlier stage disease. Surgical resection is the most effective therapy for stages I to II and selected cases of stage IIIA NSCLC²⁹. However, despite its curative intent, a high percentage of tumours will recur, with 5-year overall survival ranging from 83% for stage IA to 36% for stage IIIA disease³⁰. Adjuvant cytotoxic therapy with a cisplatin-based doublet is associated with improved survival in patients with completely resected stage II and IIIA NSCLC, with a likely benefit for stage IB disease measuring 4 cm or more³¹. Morbidity and outcomes have also improved for early stage lung cancer owing to advances in surgical and radiation technologies, including robotic and video-assisted surgery³², as well as stereotactic and hyper-fractionated approaches³³. The standard therapy for patients with unresectable locally advanced NSCLC is the combination of cytotoxic therapy and thoracic radiation, with a survival advantage for concurrent therapy over sequential approaches^{34,35}.

The high lung cancer mortality is due to the presence of metastatic disease at the time of diagnosis in most patients, indicating that improvements in long-term survival will require more effective systemic therapies^{36,37}. Molecularly targeted therapies in patients with NSCLC were initially used in the late 1990s with the introduction of gefitinib, an oral EGFR tyrosine kinase inhibitor (TKI). In unselected populations, the response rates were approximately 10%, with increased frequency of responses noted in females, never smokers, and patients of Asian descent^{38,39}. Erlotinib, another TKI against EGFR, was associated with improved survival compared to the best supportive care in patients with previously treated advanced-stage NSCLC⁴⁰. Retrospective studies subsequently demonstrated that activating *EGFR* mutations were observed in the vast majority of patients who benefited from EGFR TKIs^{41,42}. Since then, additional gene alterations, including *ALK* rearrangements, *ROS1* fusions and *BRAF* mutations led to the development of effective targeted therapies⁴³.

An important advance in the management of advanced stage NSCLC occurred in 2015, when the US FDA approved the ICB nivolumab for the treatment of patients whose disease progressed during or after platinum-based therapy, heralding a new era in the management of lung cancer⁴⁴.

Targeted therapy

The identification of targetable gene alterations has transformed the management of lung cancer, with the incorporation of tumour genotyping to allow individualized therapy and leading to remarkable responses in selected patients treated with matched TKIs (Table 1). In the multicentre Lung Cancer Mutation Consortium, targetable oncogenic drivers were observed in 64% of patients with LUAD, for whom the use of genotype-directed therapy was associated with improved survival compared to those treated without targeted therapies⁴⁵.

EGFR

EGFR belongs to a receptor tyrosine kinase family that also includes human epidermal growth factor receptor 2 (HER2, also known as ERBB2), HER3 (ERBB3) and HER4 (ERBB4). The receptor contains four extracellular domains, a transmembrane domain, a tyrosine kinase domain, and a carboxy tail⁴⁶. Binding of activating ligands leads to EGFR dimerization and *trans*-phosphorylation of the tyrosine residues in the carboxy tail, with activation of downstream pathways involved in cell proliferation, survival, invasion and angiogenesis⁴⁷. Heterozygous mutations clustering around the ATP-binding pocket of the tyrosine kinase domain may lead to constitutive EGFR activation and ligand independence^{41,42}. The most common EGFR mutations associated with sensitivity to EGFR TKIs include exon 19 deletions and a missense mutation on exon 21 (L858R)⁴⁸.

First-generation EGFR TKIs, including gefitinib and erlotinib, have shown higher objective response rates (ORRs) and progression-free survival (PFS) compared to cytotoxic therapy in previously untreated patients with EGFR mutations^{49–54}. In contrast to first-generation EGFR TKIs, which are reversible competitive ATP inhibitors that target only EGFR, second-generation inhibitors including afatinib and dacomitinib are irreversible inhibitors that also target HER2 and HER4. Both afatinib and dacomitinib showed improved PFS compared to gefitinib^{55,56}. Afatinib also demonstrated a significant improvement in the median overall survival compared to platinum-based cytotoxic therapy in patients with an exon 19 deletion, but not in those with an L858R mutation⁵⁷. Differences in outcomes between these two most common EGFR mutations may occur owing to distinct conformational changes within the ATP-binding pocket and patterns of auto-phosphorylation induced by each mutation⁵⁸.

The most common cause of acquired resistance to first-generation TKIs is a second EGFR mutation in exon 20, with a threonine-to-methionine substitution on codon 790 (T790M)^{59,60}. This mutation affects the initial EGFR TKI efficacy either from steric hindrance or by increased affinity of the tyrosine kinase domain for ATP. Other mechanisms of resistance include amplifications in HER2 or mutations in MET, BRAF or phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA), and SCLC transformation⁶¹, indicating that repeated molecular profiling at progression is needed to determine the next appropriate treatment. Third-generation EGFR TKIs are selective inhibitors of both the original sensitizing and T790M mutations, while sparing wild-type EGFR. These drugs bind covalently to cysteine on codon 797, overcoming the enhanced ATP affinity from the T790M mutation. Osimertinib, a third-generation EGFR TKI, is effective in patients with NSCLC harbouring EGFR(T790M) mutations following progression after first-generation EGFR TKI⁶², and showed increased ORRs and PFS compared to platinum-based cytotoxic therapy⁶³. In a randomized trial that compared osimertinib to either erlotinib or gefitinib in previously untreated patients with advanced stage NSCLC harbouring either EGFR exon 19 deletion or L858R mutation, osimertinib was associated with a significant improvement in PFS, establishing osimertinib as a first-line EGFR TKI option⁶⁴. Additional follow-up from this trial, including

updates on overall survival, may provide a guide to first-line therapy in previously untreated patients with tumours harbouring an EGFR mutation.

One of the mechanisms for acquired resistance to third-generation EGFR TKIs is the C797S mutation⁶⁵ which in combination with the sensitizing mutation and without T790M causes resistance to third-generation EGFR TKIs, but not to gefitinib or afatinib. The presence of triple mutants (sensitizing mutation, T790M and C797S) however, leads to resistance to all three generations of EGFR TKIs⁶⁶. Promising approaches to triple mutants include the allosteric inhibitor EAI045 in tumours with original L858R-sensitizing mutation, and brigatinib, an ALK inhibitor with activity against EGFR mutations, in tumours harbouring exon 19 deletion, both in combination with cetuximab, an anti-EGFR monoclonal antibody^{67,68}.

ALK

ALK encodes a transmembrane receptor tyrosine kinase with unclear function in humans⁶⁹. In ALK rearrangements, the most common partner is the echinoderm microtubule-associated protein-like 4 (EML4) gene (EML4-ALK)⁷⁰. Crizotinib is an oral competitive ATP inhibitor of ALK, MET and ROS1 tyrosine kinases with activity against ALK fusion-positive NSCLC⁷¹. Crizotinib is associated with improved ORRs and median PFS compared to cytotoxic therapy in both previously treated and untreated patients^{72,73}. Most patients previously treated with crizotinib benefit from second-generation ALK inhibitors including ceritinib, alectinib and brigatinib^{74–76}. Ceritinib also increased median PFS compared to first-line cytotoxic therapy in patients with ALK-positive NSCLC⁷⁷. Two randomized studies showed increased ORRs and median PFS for alectinib compared to crizotinib in patients with previously untreated ALK-positive NSCLC, establishing alectinib as a first-line treatment option^{78,79}.

Resistance to ALK inhibitors may occur owing to ALK alterations such as mutations and amplification, or upregulation of bypass signalling pathways including EGFR and mitogen-activated protein kinase (MAPK). Secondary ALK mutations are the predominant mechanism of resistance to second-generation TKIs⁸⁰. The most common ALK resistance mutation among patients treated with second-generation TKIs is G1202R, which is

Table 1 | Selected randomized trials with first-line targeted therapies

Study	Target	Design	ORR (%)	PFS (months)
IPASS ^{49,50}	EGFR	Gefitinib vs carboplatin plus paclitaxel	72.1 vs 47.3	9.5 vs 6.3
NEJ002 ⁵¹	EGFR	Gefitinib vs carboplatin plus paclitaxel	73.7 vs 30.7	10.8 vs 5.4
WJTOG-3405 ⁵²	EGFR	Gefitinib vs cisplatin plus docetaxel	62.1 vs 32.2	9.2 vs 6.3
EURTAC ⁵⁴	EGFR	Erlotinib vs platinum doublet	58 vs 15	9.7 vs 5.2
OPTIMAL ⁵³	EGFR	Erlotinib vs carboplatin plus gemcitabine	83 vs 36	13.1 vs 4.6
LUX-Lung-7 ⁵⁵	EGFR	Afatinib vs gefitinib	72.5 vs 56	11 vs 10.9
ARCHER-1050 ⁵⁶	EGFR	Dacomitinib vs gefitinib	75 vs 72	14.7 vs 9.2
FLAURA ⁶⁴	EGFR	Osimertinib vs gefitinib or erlotinib	80 vs 76	18.9 vs 10.2
PROFILE -1014 ⁷³	ALK	Crizotinib vs cisplatin plus pemetrexed	74 vs 45	10.9 vs 7
ASCEND-4 ⁷⁷	ALK	Ceritinib vs platinum plus pemetrexed	72.5 vs 26.7	16.6 vs 8.1
ALEX ⁷⁹	ALK	Alectinib vs crizotinib	82.9 vs 75.5	NR vs 11.1

The IPASS trial encompasses a subgroup of patients with EGFR mutation. NR, not reached; ORR, objective response rate; PFS, progression-free survival.

Table 2 | Randomized phase 3 clinical trials comparing ICBs to cytotoxic therapy

Study	Drug	ORR (%)	PFS (months)	OS (months)
Second-line				
CheckMate-017 (squamous) ¹¹⁰	Nivolumab 3 mg kg ⁻¹ Q2W	20	3.5	9.2
Checkmate-057 (non-squamous) ¹¹¹	Nivolumab 3 mg kg ⁻¹ Q2W	19	2.3	12.2
Keynote-010 (NSCLC PD-L1 positive) ¹¹²	Pembrolizumab 2 mg kg ⁻¹ Q3W	18	3.9	10.4
Oak (NSCLC) ¹¹³	Atezolizumab 1,200 mg Q3W	14	2.8	13.8
Combined docetaxel results* (refs 110–113)	Docetaxel 75 mg m ⁻² Q3W	9–13	2.8–4.2	6–9.6
First-line				
Keynote-024 (PD-L1 ≥ 50%) ¹¹⁴	Pembrolizumab 200 mg Q3W	44.8	10.3	NR
Checkmate-026 (PD-L1 ≥ 5%) ¹¹⁵	Nivolumab 3 mg kg ⁻¹ Q2W	26	4.2	14.4
Combined chemotherapy results† (refs 114, 115)	Platinum-based chemotherapy	27–33	5.9–6	13.2

OS, overall survival; Q2W, every two weeks; Q3W, every three weeks.

*Combined control arm results from CheckMate trials 017 and 057, Keynote-010 and Oak.

†Combined control arm results from CheckMate-026 and Keynote-024.

associated with *in vitro* resistance to all currently available ALK inhibitors except for lorlatinib, a potent third-generation ALK inhibitor with activity against most known ALK resistance mutations and efficacy in patients previously treated with up to three previous lines of ALK inhibitors⁸¹.

ROS1

ROS1 encodes a receptor tyrosine kinase that becomes constitutively activated when a rearrangement leads to the fusion of its tyrosine kinase domain with a partner gene such as *CD74*⁸². Owing to the high homology between the kinase domains of *ROS1* and ALK, drugs used to treat ALK-positive tumours including crizotinib⁸³, ceritinib⁸⁴ and lorlatinib⁸¹ have also shown marked activity in *ROS1*-positive tumours. Mechanisms of acquired resistance of *ROS1* rearrangements to crizotinib include secondary mutations, most commonly G2032R, wild-type EGFR signalling activation, *KRAS* and *KIT* mutations^{85,86}.

Other alterations

Among patients with NSCLC and *BRAF* mutation, approximately half have a single transversion at exon 15, in which valine is replaced by glutamate at residue 600 (V600E)^{87,88}, predicting for sensitivity to the *BRAF* inhibitors vemurafenib and dabrafenib as single agents^{89,90}, or dabrafenib in combination with the MEK inhibitor trametinib⁹¹.

Somatic mutations that affect *MET* exon 14, which contains the Y1003 residue required for the recruitment of CBL ubiquitin ligase that targets *MET* for ubiquitin-mediated degradation, lead to increased *MET* stability and prolonged signalling from hepatocyte growth factor stimulation⁹². Patients with NSCLC harbouring *MET* exon 14 skipping may respond to *MET* inhibitors including crizotinib or cabozantinib^{93,94}.

Other potential targetable gene alterations include mutations in *HER2*, rearrangements in the proto-oncogene *RET*, which encodes a receptor tyrosine kinase, and fusions of the neurotrophic tyrosine receptor kinase (*NTRK*) genes 1, 2 and 3, which code for tropomyosin receptor kinases (TRK) A, B and C, respectively. Initial results from targeted treatment against *HER2* and *RET* alterations have shown modest activity compared to other oncogenic targets, which may reflect their roles as dominant clonal drivers^{95–99}. By contrast, selective TRK TKIs have demonstrated histology-agnostic efficacy in patients with *NTRK* fusion-positive cancers, which occur in less than 1% of NSCLC. Although responses to TRK inhibition are notable and durable¹⁰⁰, the duration of response may eventually be limited by resistance, and a next-generation TKI has been identified to overcome acquired resistance to previous TRK kinase inhibition¹⁰¹.

Immunotherapy

Harnessing the host immune response to treat cancer is not a new concept¹⁰², and the introduction of ICBs such as monoclonal antibodies that target cytotoxic T-lymphocyte antigen-4 (CTLA-4) and antibodies against PD-1 or PD-L1 have signalled a new direction for lung cancer care. The first ICB approved by the US FDA was ipilimumab, a human immunoglobulin G1 monoclonal antibody that blocks CTLA-4, for the treatment of metastatic melanoma^{103,104}. During tumorigenesis, PD-1 signalling, driven primarily by adaptive expression of PD-L1 within the tumour, inactivates T cells that recognize tumour-specific antigens, allowing tumour progression and metastasis^{105,106}. Blocking the PD-1 and PD-L1 axis with antibodies offers an approach to restoring T cell-mediated antitumour immunity^{107–109}. ICBs have shown significant benefit in a broad population of patients with NSCLC (Table 2). Current ICBs approved or in development for NSCLC include the anti-PD-1 antibodies nivolumab (human IgG4) and pembrolizumab (humanized IgG4), as well as the anti-PD-L1 antibodies atezolizumab (human IgG1, with the Fc domain engineered to prevent antibody-directed cell cytotoxicity), durvalumab (human IgG1 engineered), and avelumab (human IgG1 demonstrating preclinical antibody-directed cell cytotoxicity activity)²⁹.

ICBs have been approved as a standard of care for patients with advanced NSCLC whose tumours progress on first-line cytotoxic therapy. Treatment with nivolumab was associated with significantly longer median overall survival compared to treatment with docetaxel among patients with metastatic NSCLC who had disease progression during or after platinum-based cytotoxic therapy^{110,111}. Subsequently, other ICBs showed improvement in overall survival compared to treatment with docetaxel, including pembrolizumab¹¹² and atezolizumab¹¹³.

In the first-line NSCLC setting, pembrolizumab was established as a new standard of care for patients with advanced or metastatic NSCLC with tumour PD-L1 expression levels of 50% or more (by a companion immunohistochemistry test), which is present in up to 30% of NSCLC. In these patients, pembrolizumab is associated with a significant improvement in ORRs, PFS and overall survival compared to platinum-based cytotoxic therapy¹¹⁴. By contrast, among patients with PD-L1 expression levels of 5% or more, nivolumab was not associated with improvements in PFS or overall survival compared to cytotoxic therapy¹¹⁵. Among patients in the nivolumab group with both PD-L1 expression levels above 50% and high TMB, the ORR was 75%, confirming previous observations that both TMB and PD-L1 expression may predict for benefit from ICB. The better toxicity profile with ICBs and equivalent survival outcome could support their selection as first-line therapy for patients with PD-L1-expressing advanced NSCLC, especially in those not eligible for cytotoxic therapy¹¹⁶.

Cytotoxic therapy could synergize with ICBs by killing tumour cells, improving the T-cell-to-tumour ratio and restoring the metabolic restrictions that result in T cell hyporesponsiveness in cancer^{117,118}. Cytotoxic therapy could also reduce immunosuppressive factors released by tumours or promote the release of antigens for presentation, broadening the antitumour T cell response. The combination of pembrolizumab, carboplatin and pemetrexed resulted in improved ORRs and PFS compared to cytotoxic therapy alone, and this combination could be an effective and tolerable first-line treatment option for patients with advanced non-LUSC¹¹⁹. Several randomized studies are evaluating the role of ICBs in combination with platinum-based cytotoxic regimens, with or without bevacizumab, in first-line advanced or recurrent NSCLC. The magnitude of the overall survival benefit of such combinations will determine their future use.

Combining anti-PD-(L)1 and anti-CTLA-4 monoclonal antibodies may result in higher and more durable responses in NSCLC, as observed in experimental models¹²⁰, and suggested in a single-arm clinical study¹²¹. Randomized studies that combine nivolumab with ipilimumab, and durvalumab with the anti-CTLA-4 human IgG4 antibody tremelimumab, are continuing and overall survival data from these studies are eagerly awaited. It is expected that the safety profile for such combinations may not be as favourable as those with ICB plus cytotoxic therapy combinations, especially in terms of immune-related adverse events. Nevertheless,

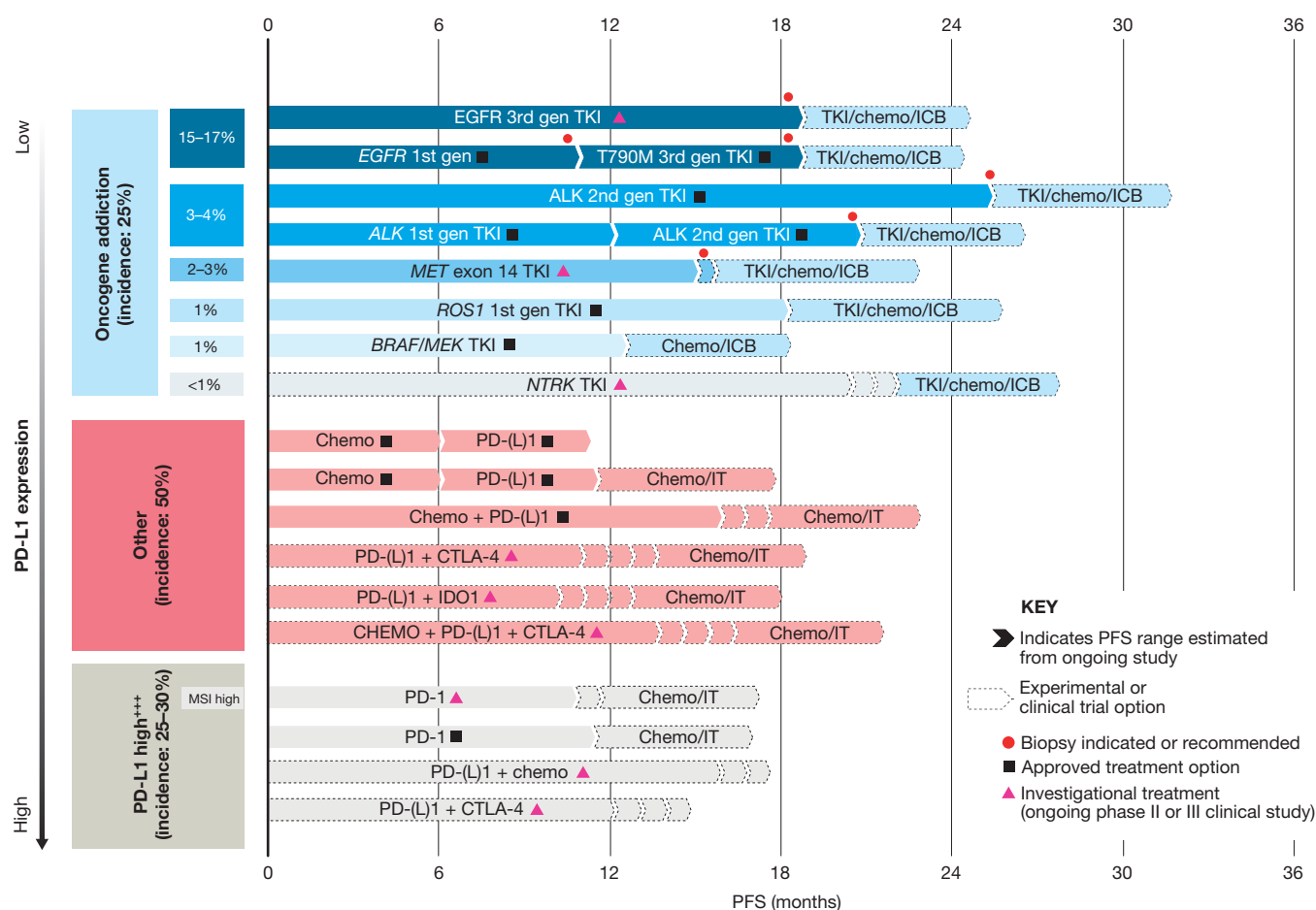


Figure 2 | Current and investigative treatment options for advanced or metastatic NSCLC. Illustration of the current and future personalized treatment options for NSCLC. Targetable oncogenic drivers account for approximately 25% of NSCLC, of which *EGFR* mutations are the most frequent⁴⁵. Biopsies are indicated at the time of disease progression to determine the best treatment option. For patients with tumours expressing high levels of PD-L1 (>50%) or high levels of microsatellite instability (MSI), single agent ICB is indicated¹¹⁴. In general, median PFS

is not the best indicator to capture the overall true benefit of ICBs, as a proportion of patients remain alive or disease-free even after long-term follow-up. In patients with tumours with high (>50%) or low (>1%) expression levels of PD-L1, current studies are assessing the benefit of anti-PD-(L)1 combinations with cytotoxic therapy, anti-CTLA-4, or other immunotherapy (IT) approaches. PFS estimates illustrated for targeted therapies from refs 52, 54, 64, 73, 79, 83, 100 and for ICBs from refs 114, 119, 121.

despite encouraging results with prolonged benefit in selected patients, most lung tumours are either inherently resistant or will adapt to or become resistant to current immunotherapies. The challenge is to develop rational combinations that will increase responses, or delay the onset of resistance¹²².

Whereas ICB monotherapy may be appropriate for tumours with high expression levels of PD-L1 or a high nonsynonymous TMB, a different approach may be required for tumours with fewer T cells and a lower TMB (Supplementary Table 1). The induction of immunogenic cancer death with the use of cytotoxic therapy, epigenetic modifiers¹²³ or oncolytic viruses¹²⁴ appears promising in preclinical models or in early phase 1 clinical studies. Another strategy includes combination with anti-angiogenic drugs, as VEGF contributes to an immunosuppressive TME by recruiting suppressive immune cells, such as myeloid-derived suppressor cells and regulatory T cells¹²⁵. Furthermore, angiogenic inhibitors may increase immune cell infiltration¹²⁶.

In tumours with a low TMB, few T cells and low PD-L1 expression ('cold tumours'), the challenge for immunotherapeutic approaches is not only to attract effector T cells into the TME, but also to present tumour antigens to T cells. Possible approaches for such cold NSCLCs could include the use of adoptive transfer of autologous tumour-infiltrating lymphocytes or chimaeric antigen receptor T cell therapy. For the latter, NSCLC-specific or unique cell-surface antigens will need to be identified. Other approaches being developed for solid tumours, and in the future

for NSCLC, include the use of autologous vaccines that use genomic information from a specific tumour to predict neo-epitopes that could be presented to T cells, for the design and manufacture of a vaccine unique for each patient¹²⁷.

Most patients who achieve an initial benefit from an ICB eventually develop resistance. Some of the mechanisms for acquired resistance to ICBs include defects in interferon- γ signalling or major histocompatibility complex presentation¹²⁸, and increased levels of the enzyme indoleamine 2,3-dioxygenase (IDO1), which catabolizes tryptophan, an amino acid required for optimal T cell function^{129,130}.

ICBs are poised to move to earlier stages of lung cancer therapy, in an attempt to improve survival after surgery or radiotherapy, in which the goal is curative. A randomized trial of durvalumab as a sequential treatment in patients with locally advanced unresectable stage NSCLC who had not progressed after standard concurrent platinum-based cytotoxic and radiation therapy showed a notable improvement in median PFS compared to placebo¹³¹. The role of ICBs in patients with curable non-metastatic disease will be further clarified with the results of many continuing trials accruing in both the perioperative setting and in patients with locally advanced disease treated with concurrent cytotoxic and radiation therapy.

The notable clinical success of cancer immunotherapy over a short period of time suggests that it may form the foundation of future curative-intent regimens for many malignancies, including NSCLC (Fig. 2).

Advancing personalized medicine and trial design

The rapid development of targeted therapies with newer and more potent generations of drugs, and the characterization of the mechanisms of acquired resistance, established the role for repeated genomic profiling at the time of tumour progression, particularly in patients with *EGFR* mutations, in which the most common cause of resistance, the T790M mutation, may be successfully treated with osimertinib. The repeated genomic profile has traditionally been performed through repeated biopsy, which may not always be feasible and carries risks for complications. An emerging option is the use of plasma genotyping with sequencing circulating tumour DNA (ctDNA). In a retrospective analysis including 216 patients with both central tissue and plasma genotyping available before treatment with osimertinib in a clinical trial, the sensitivity of plasma genotyping for *EGFR*(T790M) was 70%¹³². Furthermore, plasma *EGFR*(T790M) mutations were found in 31% of patients with tissue-negative results. Because the outcomes with osimertinib treatment, including ORRs and PFS, were similar for T790M mutations detected by tissue or plasma, a repeated biopsy could be avoided in those with positive plasma results. By contrast, patients with negative plasma results should undergo a repeated biopsy. Multianalytical ctDNA tests for multiple molecular markers, including the detection of high TMB, are being developed to support individualized lung cancer treatment.

Because many gene alterations are uncommon or rare, and hundreds of combination studies are being conducted in NSCLC, recruiting patients to clinical trials is becoming more difficult. To adjust to this current situation, innovative master protocols are being used in which several questions can be answered in one study. Such master protocols are designed to encompass a collection of trials that share key design features and used to assess either multiple targeted therapies or immunotherapy combinations for a single disease (umbrella trials) or a single targeted therapy for multiple diseases (basket trials). The Biomarker-integrated Approaches of Targeted Therapy for Lung cancer Elimination (BATTLE) and Lung Cancer Master Protocol (Lung-MAP) trials are among the first umbrella trials for patients with NSCLC, with the former showing the feasibility of performing fresh biopsies to guide the next line of therapy, a principle that has been often used in the most recent studies^{133,134}.

Future perspectives

The treatment of NSCLC has undergone remarkable changes. A better understanding of the tumour biology enabled the development of targeted therapies that heralded the era of personalized medicine. Furthermore, the introduction of ICBs has led to prolonged survival in selected patients. Nevertheless, this unprecedented benefit from current standard therapies is still observed in only a minority of patients, with targeted therapies restricted to non-LUSC histological subtypes that contain actionable driver mutations, and durable responses from immunotherapy occurring uncommonly. One of the main concerns with targeted therapy is the emergence of secondary clones that may not be effectively targeted by the initial treatment directed at the founder clone. To improve outcomes further, there is a need to understand better the mechanisms of acquired resistance to allow their prevention or effective treatment at the time of emergence. Therefore, focusing on both dominant and sub-clones may be required for a more effective and durable benefit from targeted therapies¹³⁵.

The application of ctDNA to track the evolutionary dynamics of early stage lung cancer should be expanded to detect both oncogenic drivers and track resistant mutations, providing a future approach for ctDNA-driven targeted therapeutics¹³⁶. A systematic approach to collect tissue samples not only at the time of diagnosis but also serially at the times of relapse to evaluate the dynamic clonal evolution that occurs over time and possibly at different metastatic sites will be crucial for further therapeutic advances.

Better predictors for response to immunotherapy are critical for its optimal use. Although both PD-L1 and TMB may be used to select patients for therapy¹³⁷, most patients will not fit the ideal profile based on these two biomarkers. Furthermore, correlating cancer genomic

information with the cellular components of the TME could allow the use of rationally designed combination therapies¹³⁸.

Over the past 20 years, there has been enormous progress in the understanding of the biology and management of NSCLC, with targeted and immunotherapies providing a new foundation towards rationally designed therapeutic regimens with manageable toxicity profiles and improvement in survival. Although the treatment of patients with metastatic NSCLC has long been considered palliative, continuing drug development provides the hope for prolonged survival in an increasing number of patients and, for the first time, raises the possibility of a cure in those with metastatic disease. Furthermore, the use of new therapeutic modalities including immunotherapy may have an even higher impact in patients with earlier non-detectable metastatic disease, in whom treatment is given with curative intent.

Received 18 August; accepted 29 November 2017.

1. Torre, L. A. *et al.* Global cancer statistics, 2012. *CA Cancer J. Clin.* **65**, 87–108 (2015).
2. Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E. & Adjei, A. A. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin. Proc.* **83**, 584–594 (2008).
3. Alberg, A. J., Brock, M. V., Ford, J. G., Samet, J. M. & Spivack, S. D. Epidemiology of lung cancer: diagnosis and management of lung cancer, 3rd ed.: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* **143**, e1S–e29S (2013).
4. Sun, S., Schiller, J. H. & Gazdar, A. F. Lung cancer in never smokers—a different disease. *Nat. Rev. Cancer* **7**, 778–790 (2007).
5. Vineis, P. *et al.* Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study. *Br. Med. J.* **330**, 277 (2005).
6. Hackshaw, A. K., Law, M. R. & Wald, N. J. The accumulated evidence on lung cancer and environmental tobacco smoke. *Br. Med. J.* **315**, 980–988 (1997).
7. Shahab, L. *et al.* Nicotine, carcinogen, and toxin exposure in long-term e-cigarette and nicotine replacement therapy users: a cross-sectional study. *Ann. Intern. Med.* **166**, 390–400 (2017).
8. Hays, J. T. & Ebbert, J. O. Varenicline for tobacco dependence. *N. Engl. J. Med.* **359**, 2018–2024 (2008).
9. Brandon, T. H. *et al.* Electronic nicotine delivery systems: a policy statement from the American Association for Cancer Research and the American Society of Clinical Oncology. *J. Clin. Oncol.* **33**, 952–963 (2015).
10. Zhang, J. *et al.* Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346**, 256–259 (2014).
11. Engelman, J. A. *et al.* MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* **316**, 1039–1043 (2007).
12. Turke, A. B. *et al.* Preexistence and clonal selection of MET amplification in *EGFR* mutant NSCLC. *Cancer Cell* **17**, 77–88 (2010).
13. Pao, W. *et al.* KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Med.* **2**, e17 (2005).
14. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075 (2008).
15. Reck, M. & Rabe, K. F. Precision diagnosis and treatment for advanced non-small-cell lung cancer. *N. Engl. J. Med.* **377**, 849–861 (2017).
16. Ahrendt, S. A. *et al.* p53 mutations and survival in stage I non-small-cell lung cancer: results of a prospective study. *J. Natl. Cancer Inst.* **95**, 961–970 (2003).
17. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
18. Robinson, D. R. *et al.* Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297–303 (2017).
19. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351**, 1463–1469 (2016).
20. Le, D. T. *et al.* PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
21. Koyama, S. *et al.* STK11/LKB1 deficiency promotes neutrophil recruitment and proinflammatory cytokine production to suppress T-cell activity in the lung tumor microenvironment. *Cancer Res.* **76**, 999–1008 (2016).
22. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012). **References 17 and 22 are landmark genomics analyses that describe the molecular landscapes of lung adenocarcinoma and squamous cell carcinoma, respectively.**
23. Hanna, N. *et al.* Systemic therapy for stage IV non-small-cell lung cancer: american society of clinical oncology clinical practice guideline update. *J. Clin. Oncol.* **35**, 3484–3515 (2017).
24. Schiller, J. H. *et al.* Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N. Engl. J. Med.* **346**, 92–98 (2002).
25. Scagliotti, G. V. *et al.* Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naïve patients with advanced-stage non-small-cell lung cancer. *J. Clin. Oncol.* **26**, 3543–3551 (2008).
26. Sandler, A. *et al.* Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N. Engl. J. Med.* **355**, 2542–2550 (2006).

27. Thatcher, N. *et al.* Necitumumab plus gemcitabine and cisplatin versus gemcitabine and cisplatin alone as first-line therapy in patients with stage IV squamous non-small-cell lung cancer (SQUIRE): an open-label, randomised, controlled phase 3 trial. *Lancet Oncol.* **16**, 763–774 (2015).
28. Garon, E. B. *et al.* Ramucicirumab plus docetaxel versus placebo plus docetaxel for second-line treatment of stage IV non-small-cell lung cancer after disease progression on platinum-based therapy (REVEL): a multicentre, double-blind, randomised phase 3 trial. *Lancet* **384**, 665–673 (2014).
29. Hirsch, F. R. *et al.* Lung cancer: current therapies and new targeted treatments. *Lancet* **389**, 299–311 (2017).
30. Goldstraw, P. *et al.* The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J. Thorac. Oncol.* **11**, 39–51 (2016).
31. Kris, M. G. *et al.* Adjuvant systemic therapy and adjuvant radiation therapy for stage I to IIIA completely resected non-small-cell lung cancers: American Society of Clinical Oncology/Cancer Care Ontario clinical practice guideline update. *J. Clin. Oncol.* **35**, 2960–2974 (2017).
32. Liang, H. *et al.* Robotic versus video-assisted lobectomy/segmentectomy for lung cancer: a meta-analysis. *Ann. Surg.* <https://doi.org/10.1097/SLA.0000000000002346> (2017).
33. Chi, A., Chen, H., Wen, S., Yan, H. & Liao, Z. Comparison of particle beam therapy and stereotactic body radiotherapy for early stage non-small cell lung cancer: A systematic review and hypothesis-generating meta-analysis. *Radiother. Oncol.* **123**, 346–354 (2017).
34. Dillman, R. O. *et al.* A randomized trial of induction chemotherapy plus high-dose radiation versus radiation alone in stage III non-small-cell lung cancer. *N. Engl. J. Med.* **323**, 940–945 (1990).
35. Curran, W. J., Jr *et al.* Sequential vs. concurrent chemoradiation for stage III non-small cell lung cancer: randomized phase III trial RTOG 9410. *J. Natl. Cancer Inst.* **103**, 1452–1460 (2011).
36. Herbst, R. S., Heymach, J. V. & Lippman, S. M. Lung cancer. *N. Engl. J. Med.* **359**, 1367–1380 (2008).
37. Morgensztern, D., Ng, S. H., Gao, F. & Govindan, R. Trends in stage distribution for patients with non-small cell lung cancer: a National Cancer Database survey. *J. Thorac. Oncol.* **5**, 29–33 (2010).
38. Fukuoka, M. *et al.* Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial) [corrected]. *J. Clin. Oncol.* **21**, 2237–2246 (2003).
39. Kris, M. G. *et al.* Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. *J. Am. Med. Assoc.* **290**, 2149–2158 (2003).
40. Shepherd, F. A. *et al.* Erlotinib in previously treated non-small-cell lung cancer. *N. Engl. J. Med.* **353**, 123–132 (2005).
41. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004).
42. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500 (2004).
- References 41 and 42 were among the first studies to demonstrate that EGFR mutations in NSCLC confer sensitivity to anti-EGFR tyrosine kinase inhibitors.**
43. Rikova, K. *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**, 1190–1203 (2007).
44. Kazandjian, D. *et al.* FDA approval summary: nivolumab for the treatment of metastatic non-small cell lung cancer with progression on or after platinum-based chemotherapy. *Oncologist* **21**, 634–642 (2016).
45. Kris, M. G. *et al.* Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *J. Am. Med. Assoc.* **311**, 1998–2006 (2014).
46. Lemmon, M. A., Schlessinger, J. & Ferguson, K. M. The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harb. Perspect. Biol.* **6**, a020768 (2014).
47. Wheeler, D. L., Dunn, E. F. & Harari, P. M. Understanding resistance to EGFR inhibitors-impact on future treatment strategies. *Nat. Rev. Clin. Oncol.* **7**, 493–507 (2010).
48. Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **7**, 169–181 (2007).
49. Mok, T. S. *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N. Engl. J. Med.* **361**, 947–957 (2009).
- This study relates to a change in the era of personalized therapy, and demonstrates that an anti-EGFR tyrosine kinase inhibitor is superior to cytotoxic therapy in patients with tumours that contain an activating EGFR mutation.**
50. Fukuoka, M. *et al.* Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients with advanced non-small-cell lung cancer in Asia (IPASS). *J. Clin. Oncol.* **29**, 2866–2874 (2011).
51. Maemondo, M. *et al.* Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N. Engl. J. Med.* **362**, 2380–2388 (2010).
52. Mitsudomi, T. *et al.* Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. *Lancet Oncol.* **11**, 121–128 (2010).
53. Zhou, C. *et al.* Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol.* **12**, 735–742 (2011).
54. Rosell, R. *et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* **13**, 239–246 (2012).
55. Paz-Ares, L. *et al.* Afatinib versus gefitinib in patients with EGFR mutation-positive advanced non-small-cell lung cancer: overall survival data from the phase IIb LUX-Lung 7 trial. *Ann. Oncol.* **28**, 270–277 (2017).
56. Wu, Y. L. *et al.* Dacomitinib versus gefitinib as first-line treatment for patients with EGFR-mutation-positive non-small-cell lung cancer (ARCHER 1050): a randomised, open-label, phase 3 trial. *Lancet Oncol.* **18**, 1454–1466 (2017).
57. Yang, J. C. *et al.* Afatinib versus cisplatin-based chemotherapy for EGFR mutation-positive lung adenocarcinoma (LUX-Lung 3 and LUX-Lung 6): analysis of overall survival data from two randomised, phase 3 trials. *Lancet Oncol.* **16**, 141–151 (2015).
58. Okabe, T. *et al.* Differential constitutive activation of the epidermal growth factor receptor in non-small cell lung cancer cells bearing EGFR gene mutation and amplification. *Cancer Res.* **67**, 2046–2053 (2007).
59. Kobayashi, S. *et al.* EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **352**, 786–792 (2005).
60. Sequist, L. V. *et al.* Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci. Transl. Med.* **3**, 75ra26 (2011).
61. Camidge, D. R., Pao, W. & Sequist, L. V. Acquired resistance to TKIs in solid tumours: learning from lung cancer. *Nat. Rev. Clin. Oncol.* **11**, 473–481 (2014).
62. Jänne, P. A. *et al.* AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N. Engl. J. Med.* **372**, 1689–1699 (2015).
63. Mok, T. S. *et al.* Osimertinib or platinum-pemetrexed in EGFR T790M-positive lung cancer. *N. Engl. J. Med.* **376**, 629–640 (2017).
64. Soria, J. C. *et al.* Osimertinib in untreated EGFR-mutated advanced non-small cell lung cancer. *N. Engl. J. Med.* <http://doi.org/10.1056/NEJMoa1713137> (2017).
65. Thress, K. S. *et al.* Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat. Med.* **21**, 560–562 (2015).
66. Niederst, M. J. *et al.* The allelic context of the C797S mutation acquired upon treatment with third-generation EGFR inhibitors impacts sensitivity to subsequent treatment strategies. *Clin. Cancer Res.* **21**, 3924–3933 (2015).
67. Jia, Y. *et al.* Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. *Nature* **534**, 129–132 (2016).
68. Uchibori, K. *et al.* Brigatinib combined with anti-EGFR antibody overcomes osimertinib resistance in EGFR-mutated non-small-cell lung cancer. *Nat. Commun.* **8**, 14768 (2017).
69. Lin, J. J., Riely, G. J. & Shaw, A. T. Targeting ALK: precision medicine takes on drug resistance. *Cancer Discov.* **7**, 137–155 (2017).
70. Soda, M. *et al.* Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566 (2007).
- This study describes the discovery of ALK rearrangements in NSCLC.**
71. Kwak, E. L. *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N. Engl. J. Med.* **363**, 1693–1703 (2010).
- This study is the first to report the activity of crizotinib in patients with ALK rearrangements.**
72. Shaw, A. T. *et al.* Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N. Engl. J. Med.* **368**, 2385–2394 (2013).
73. Solomon, B. J. *et al.* First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *N. Engl. J. Med.* **371**, 2167–2177 (2014).
74. Shaw, A. T. *et al.* Ceritinib in ALK-rearranged non-small-cell lung cancer. *N. Engl. J. Med.* **370**, 1189–1197 (2014).
75. Shaw, A. T. *et al.* Alectinib in ALK-positive, crizotinib-resistant, non-small-cell lung cancer: a single-group, multicentre, phase 2 trial. *Lancet Oncol.* **17**, 234–242 (2016).
76. Kim, D. W. *et al.* Brigatinib in patients with crizotinib-refractory anaplastic lymphoma kinase-positive non-small-cell lung cancer: a randomized, multicenter phase II trial. *J. Clin. Oncol.* **35**, 2490–2498 (2017).
77. Soria, J. C. *et al.* First-line ceritinib versus platinum-based chemotherapy in advanced ALK-rearranged non-small-cell lung cancer (ASCEND-4): a randomised, open-label, phase 3 study. *Lancet* **389**, 917–929 (2017).
78. Hida, T. *et al.* Alectinib versus crizotinib in patients with ALK-positive non-small-cell lung cancer (J-ALEX): an open-label, randomised phase 3 trial. *Lancet* **390**, 29–39 (2017).
79. Peters, S. *et al.* Alectinib versus crizotinib in untreated ALK-positive non-small-cell lung cancer. *N. Engl. J. Med.* **377**, 829–838 (2017).
80. Gainor, J. F. *et al.* Molecular mechanisms of resistance to first- and second-generation ALK inhibitors in ALK-rearranged lung cancer. *Cancer Discov.* **6**, 1118–1133 (2016).
81. Shaw, A. T. *et al.* Lorlatinib in non-small-cell lung cancer with ALK or ROS1 rearrangement: an international, multicentre, open-label, single-arm first-in-man phase 1 trial. *Lancet Oncol.* **18**, 1590–1599 (2017).
82. Facchinetti, F. *et al.* Oncogene addiction in non-small cell lung cancer: focus on ROS1 inhibition. *Cancer Treat. Rev.* **55**, 83–95 (2017).
83. Shaw, A. T. *et al.* Crizotinib in ROS1-rearranged non-small-cell lung cancer. *N. Engl. J. Med.* **371**, 1963–1971 (2014).
84. Lim, S. M. *et al.* Open-label, multicenter, phase II study of ceritinib in patients with non-small-cell lung cancer harboring Ros1 rearrangement. *J. Clin. Oncol.* **35**, 2613–2618 (2017).
85. Awad, M. M. *et al.* Acquired resistance to crizotinib from a mutation in CD74-ROS1. *N. Engl. J. Med.* **368**, 2395–2401 (2013).

86. Davies, K. D. *et al.* Resistance to ROS1 inhibition mediated by EGFR pathway activation in non-small cell lung cancer. *PLoS One* **8**, e82236 (2013).
 87. Marchetti, A. *et al.* Clinical features and outcome of patients with non-small-cell lung cancer harboring BRAF mutations. *J. Clin. Oncol.* **29**, 3574–3579 (2011).
 88. Cardarella, S. *et al.* Clinical, pathologic, and biologic features associated with BRAF mutations in non-small cell lung cancer. *Clin. Can. Res.* **19**, 4532–4540 (2013).
 89. Hyman, D. M. *et al.* Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. *N. Engl. J. Med.* **373**, 726–736 (2015).
 90. Planchard, D. *et al.* Dabrafenib plus trametinib in patients with BRAF(V600E)-positive advanced non-small-cell lung cancer: a single-arm, multicentre, open-label, phase 2 trial. *Lancet Oncol.* **17**, 642–650 (2016).
 91. Planchard, D. *et al.* Dabrafenib plus trametinib in patients with previously treated BRAF(V600E)-mutant metastatic non-small cell lung cancer: an open-label, multicentre phase 2 trial. *Lancet Oncol.* **17**, 984–993 (2016).
 92. Frampton, G. M. *et al.* Activation of MET via diverse exon 14 splicing alterations occurs in multiple tumor types and confers clinical sensitivity to MET inhibitors. *Cancer Discov.* **5**, 850–859 (2015).
 93. Paik, P. K. *et al.* Response to MET inhibitors in patients with stage IV lung adenocarcinomas harboring MET mutations causing exon 14 skipping. *Cancer Discov.* **5**, 842–849 (2015).
 94. Awad, M. M. *et al.* MET exon 14 mutations in non-small-cell lung cancer are associated with advanced age and stage-dependent MET genomic amplification and c-Met overexpression. *J. Clin. Oncol.* **34**, 721–730 (2016).
 95. Mazières, J. *et al.* Lung cancer that harbors an HER2 mutation: epidemiologic characteristics and therapeutic perspectives. *J. Clin. Oncol.* **31**, 1997–2003 (2013).
 96. Mazières, J. *et al.* Lung cancer patients with HER2 mutations treated with chemotherapy and HER2-targeted drugs: results from the European EUHER2 cohort. *Ann. Oncol.* **27**, 281–286 (2016).
 97. Kohno, T. *et al.* KIF5B-RET fusions in lung adenocarcinoma. *Nat. Med.* **18**, 375–377 (2012).
 98. Drilon, A. *et al.* Cabozantinib in patients with advanced RET-rearranged non-small-cell lung cancer: an open-label, single-centre, phase 2, single-arm trial. *Lancet Oncol.* **17**, 1653–1660 (2016).
 99. Gautschi, O. *et al.* Targeting RET in patients with RET-rearranged lung cancers: results from the global, multicenter RET registry. *J. Clin. Oncol.* **35**, 1403–1410 (2017).
 100. Hyman, D. M. *et al.* The efficacy of larotrectinib (LOXO-101), a selective tropomyosin receptor kinase (TRK) inhibitor, in adult and pediatric TRK fusion cancers. *J. Clin. Oncol.* **35**, LBA2501–LBA2501 (2017).
 101. Drilon, A. *et al.* A next-generation TRK kinase inhibitor overcomes acquired resistance to prior TRK kinase inhibition in patients with TRK fusion-positive solid tumors. *Cancer Discov.* **7**, 963–972 (2017).
 102. Coley, W. B. The treatment of malignant tumors by repeated inoculations of erysipelas. With a report of ten original cases. 1893. *Clin. Orthop. Relat. Res.* (262):3–11 (1991).
 103. Leach, D. R., Krummel, M. F. & Allison, J. P. Enhancement of antitumor immunity by CTLA-4 blockade. *Science* **271**, 1734–1736 (1996).
 104. Hodi, F. S. *et al.* Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* **363**, 711–723 (2010).
 105. Dong, H. *et al.* Tumor-associated B7-H1 promotes T-cell apoptosis: a potential mechanism of immune evasion. *Nat. Med.* **8**, 793–800 (2002).
 106. Iwai, Y. *et al.* Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. *Proc. Natl Acad. Sci. USA* **99**, 12293–12297 (2002).
 107. Topalian, S. L. *et al.* Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N. Engl. J. Med.* **366**, 2443–2454 (2012).
 108. Garon, E. B. *et al.* Pembrolizumab for the treatment of non-small-cell lung cancer. *N. Engl. J. Med.* **372**, 2018–2028 (2015).
 109. Herbst, R. S. *et al.* Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* **515**, 563–567 (2014).
 110. Brahmer, J. *et al.* Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N. Engl. J. Med.* **373**, 123–135 (2015).
 111. Borghaei, H. *et al.* Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N. Engl. J. Med.* **373**, 1627–1639 (2015).
- References 110 and 111 were the first phase 3 studies to show increased survival for ICBs compared to cytotoxic therapy in patients with previously treated advanced-stage NSCLC, heralding the era of immunotherapy for NSCLC.**
112. Herbst, R. S. *et al.* Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet* **387**, 1540–1550 (2016).
 113. Rittmeyer, A. *et al.* Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet* **389**, 255–265 (2017).
 114. Reck, M. *et al.* Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N. Engl. J. Med.* **375**, 1823–1833 (2016).
- This study provides evidence that in selected patients with high tumour expression of PD-L1, ICBs are more effective than cytotoxic therapy in the first-line setting.**
115. Carbone, D. P. *et al.* First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. *N. Engl. J. Med.* **376**, 2415–2426 (2017).
 116. Herbst, R. S. & Sznol, M. Diminished but not dead: chemotherapy for the treatment of NSCLC. *Lancet Oncol.* **17**, 1464–1465 (2016).
 117. Chang, C. H. *et al.* Metabolic competition in the tumor microenvironment is a driver of cancer progression. *Cell* **162**, 1229–1241 (2015).
 118. Zitvogel, L., Galluzzi, L., Smyth, M. J. & Kroemer, G. Mechanism of action of conventional and targeted anticancer therapies: reinstating immunosurveillance. *Immunity* **39**, 74–88 (2013).
 119. Langer, C. J. *et al.* Carboplatin and pemetrexed with or without pembrolizumab for advanced, non-squamous non-small-cell lung cancer: a randomised, phase 2 cohort of the open-label KEYNOTE-021 study. *Lancet Oncol.* **17**, 1497–1508 (2016).
 120. Curran, M. A., Montalvo, W., Yagita, H. & Allison, J. P. PD-1 and CTLA-4 combination blockade expands infiltrating T cells and reduces regulatory T and myeloid cells within B16 melanoma tumors. *Proc. Natl Acad. Sci. USA* **107**, 4275–4280 (2010).
 121. Hellmann, M. D. *et al.* Nivolumab plus ipilimumab as first-line treatment for advanced non-small-cell lung cancer (CheckMate 012): results of an open-label, phase 1, multicohort study. *Lancet Oncol.* **18**, 31–41 (2017).
 122. Sharma, P., Hu-Lieskovan, S., Wargo, J. A. & Ribas, A. Primary, adaptive, and acquired resistance to cancer immunotherapy. *Nat. Med.* **23**, 1362–1368 (2017).
 123. Bezu, L. *et al.* Combinatorial strategies for the induction of immunogenic cell death. *Front. Immunol.* **6**, 187 (2015).
 124. Lawler, S. E., Speranza, M. C., Cho, C. F. & Chiocca, E. A. Oncolytic viruses in cancer treatment: a review. *JAMA Oncol.* **3**, 841–849 (2017).
 125. Voron, T. *et al.* Control of the immune response by pro-angiogenic factors. *Front. Oncol.* **4**, 70 (2014).
 126. Tian, L. *et al.* Mutual regulation of tumour vessel normalization and immunostimulatory reprogramming. *Nature* **544**, 250–254 (2017).
 127. Sahin, U. *et al.* Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547**, 222–226 (2017).
 128. Gettinger, S. *et al.* Impaired HLA class I antigen processing and presentation as a mechanism of acquired resistance to immune checkpoint inhibitors in lung cancer. *Cancer Discov.* **7**, 1420–1435 (2017).
 129. Zaretsky, J. M. *et al.* Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N. Engl. J. Med.* **375**, 819–829 (2016).
 130. Holmggaard, R. B., Zamarin, D., Munn, D. H., Wolchok, J. D. & Allison, J. P. Indoleamine 2,3-dioxygenase is a critical resistance mechanism in antitumor T cell immunotherapy targeting CTLA-4. *J. Exp. Med.* **210**, 1389–1402 (2013).
 131. Antonia, S. J. *et al.* Durvalumab after chemoradiotherapy in stage III non-small-cell lung cancer. *N. Engl. J. Med.* **377**, 1919–1929 (2017).
 132. Oxnard, G. R. *et al.* Association between plasma genotyping and outcomes of treatment with osimertinib (AZD9291) in advanced non-small-cell lung cancer. *J. Clin. Oncol.* **34**, 3375–3382 (2016).
 133. Kim, E. S. *et al.* The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov.* **1**, 44–53 (2011).
 134. Herbst, R. S. *et al.* Lung Master Protocol (Lung-MAP)—A biomarker-driven protocol for accelerating development of therapies for squamous cell lung cancer: SWOG S1400. *Clin. Cancer Res.* **21**, 1514–1524 (2015).
 135. Blakely, C. M. *et al.* Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers. *Nat. Genet.* **49**, 1693–1704 (2017).
 136. Abbosh, C. *et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446–451 (2017).
- This study introduces ctDNA profiling to track the subclonal nature of lung cancer progression, providing an approach for ctDNA-driven therapeutic studies.**
137. Rizvi, N. A. *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
- This is a landmark study indicating that lung cancers with high non-synonymous mutation burden are more responsive to ICB.**
138. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**, 321–330 (2017).
 139. Romero, R. *et al.* Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis. *Nat. Med.* **23**, 1362–1368 (2017).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank L. Chen and A. M. Incassati for editorial assistance. R. Herbst is supported by the Yale SPOR in Lung Cancer (P50CA196530).

Author Contributions All authors contributed to the writing of this Review.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to R.S.H. (roy.herbst@yale.edu) or C.B. (chris.boshoff@pfizer.com).

Midbrain circuits that set locomotor speed and gait selection

V. Caggiano^{1†*}, R. Leiras^{1,2*}, H. Goñi-Errro^{1,2}, D. Masini³, C. Bellardita^{1,2}, J. Bouvier^{1†}, V. Caldeira¹, G. Fisone³ & O. Kiehn^{1,2}

Locomotion is a fundamental motor function common to the animal kingdom. It is implemented episodically and adapted to behavioural needs, including exploration, which requires slow locomotion, and escape behaviour, which necessitates faster speeds. The control of these functions originates in brainstem structures, although the neuronal substrate(s) that support them have not yet been elucidated. Here we show in mice that speed and gait selection are controlled by glutamatergic excitatory neurons (GlutNs) segregated in two distinct midbrain nuclei: the cuneiform nucleus (CnF) and the pedunculopontine nucleus (PPN). GlutNs in both of these regions contribute to the control of slower, alternating-gait locomotion, whereas only GlutNs in the CnF are able to elicit high-speed, synchronous-gait locomotion. Additionally, both the activation dynamics and the input and output connectivity matrices of GlutNs in the PPN and the CnF support explorative and escape locomotion, respectively. Our results identify two regions in the midbrain that act in conjunction to select context-dependent locomotor behaviours.

Activities such as exploring the surroundings, searching for food or escaping from danger depend on locomotor movements. The episodic nature of locomotion requires cycles of initiation and termination. In addition, during locomotion and depending on behavioural demands, changes of speed are necessary. In quadrupeds this function is often associated with changes in limb coordination, resulting in different gaits¹. In mice, the alternating gaits—walk and trot—are associated with slow locomotor speeds, whereas the synchronous gaits—gallop and bound—involve fast locomotor speeds¹ and are mostly used during escape-like behaviour. The executive locomotor circuits that control the coordination of muscle activity are localized in the spinal cord^{2–6}, however the commands for initiation and gait selection may originate in different supraspinal structures. The most important neuronal structure that has been implicated in these functions is the mesencephalic locomotor region (MLR)^{7–9}, which is located in the midbrain.

The MLR was first defined functionally in cats as a region localized in or around the CnF, in which continuous electrical stimulation evoked persistent locomotion¹⁰. Analogues of the MLR have been observed in many vertebrates—including fish, rodents, primates and humans^{8,9,11,12}—but with conflicting results as to their anatomical location. In addition to the CnF, the more ventrally located PPN has also been implicated in locomotor control. Besides being anatomically separated, each of these regions contain neurons with diverse transmitter phenotypes with excitatory long-range projection neurons—glutamatergic in the CnF and both glutamatergic and cholinergic in the PPN—intermingled with local inhibitory interneurons^{11,12}. Electrical stimulation or lesion studies are therefore unable to distinguish the contribution from the various intermingled neuronal populations present in these areas^{12,13}. Recently, optogenetic manipulations have shown that stimulation of GlutNs in and around the PPN induces locomotion in mice¹⁴. The MLR has thus been previously regarded as a single entity, precluding any evaluation of the putative divergent control of locomotion by subpopulations of neurons in the CnF and the PPN. As such, the question of whether—and if so, how—neuronal

populations of the CnF and the PPN control locomotion remains unanswered.

Here we address this question by using cell-type-specific targeting to modulate and record the activity of neurotransmitter-defined neurons in either the CnF or the PPN. Our results reveal that the MLR is defined by glutamatergic subpopulations of neurons in both the PPN and the CnF that may act in conjunction to control slower, alternating-gait locomotion. Furthermore, glutamatergic neurons in the PPN promote locomotion for the purpose of explorative behaviour, whereas those in the CnF promote escape locomotion. Our study identifies circuits that have key roles in the appropriate command pathways for selecting locomotor outputs contingent on behavioural contexts.

Control of speed by CnF and PPN cells

The anatomical locations of the CnF and the PPN are shown in Fig. 1a, b. The glutamatergic cells in the CnF and the PPN express the vesicular glutamate transporter 2, Vglut2 (Allen Brain Atlas and ref. 15) (Fig. 1c). Therefore, to target glutamatergic neurons in the CnF or the PPN, we injected a Cre-dependent adeno-associated virus (AAV) carrying channelrhodopsin-2 (ChR2) and the fluorescent tags mCherry or enhanced yellow fluorescent protein (eYFP) (denoted AAV-DIO-ChR2-eYFP/mCherry) into *Vglut2^{Cre}* mice (ref. 16, Fig. 1d, f; injection sites in Extended Data Fig. 1a, b).

In a linear corridor¹, unilateral light activation of Vglut2⁺ChR2 CnF neurons led to the initiation of full-body locomotion in resting mice ($N = 9$ out of 9 mice; locomotor movements detected in 115 out of 131 trials, 88%). Increasing the stimulation frequencies stepwise from threshold values of around 2–5 Hz to maximum frequencies at 50 Hz progressively increased the speed of locomotion ($P < 0.05$, Kruskal–Wallis test, post hoc analysis with Bonferroni correction for multiple comparisons; between speeds at different frequencies shown in Fig. 1e, h, blue line and Extended Data Fig. 2a; $P < 0.001$, Spearman correlation $r = 0.32$ between frequency of stimulation and maximum speed, Supplementary Video 1). The activation of Vglut2⁺ChR2

¹Mammalian Locomotor Laboratory, Department of Neuroscience, Karolinska Institutet, 17177 Stockholm, Sweden. ²Department of Neuroscience, University of Copenhagen, Blegdamsvej 3, 2200 Copenhagen, Denmark. ³Laboratory of Molecular Neuropharmacology, Department of Neuroscience, Karolinska Institutet, 17177 Stockholm, Sweden. [†]Present addresses: Computational Biology Center, IBM T.J. Watson Research Center, 1101 Kitchawan Road, Route 134, Yorktown Heights, New York 10598, USA (V.Cag.); Paris-Saclay Institute of Neuroscience, UMR9197, CNRS and Université Paris-11, 91190 Gif-sur-Yvette, France (J.B.).

*These authors contributed equally to this work.

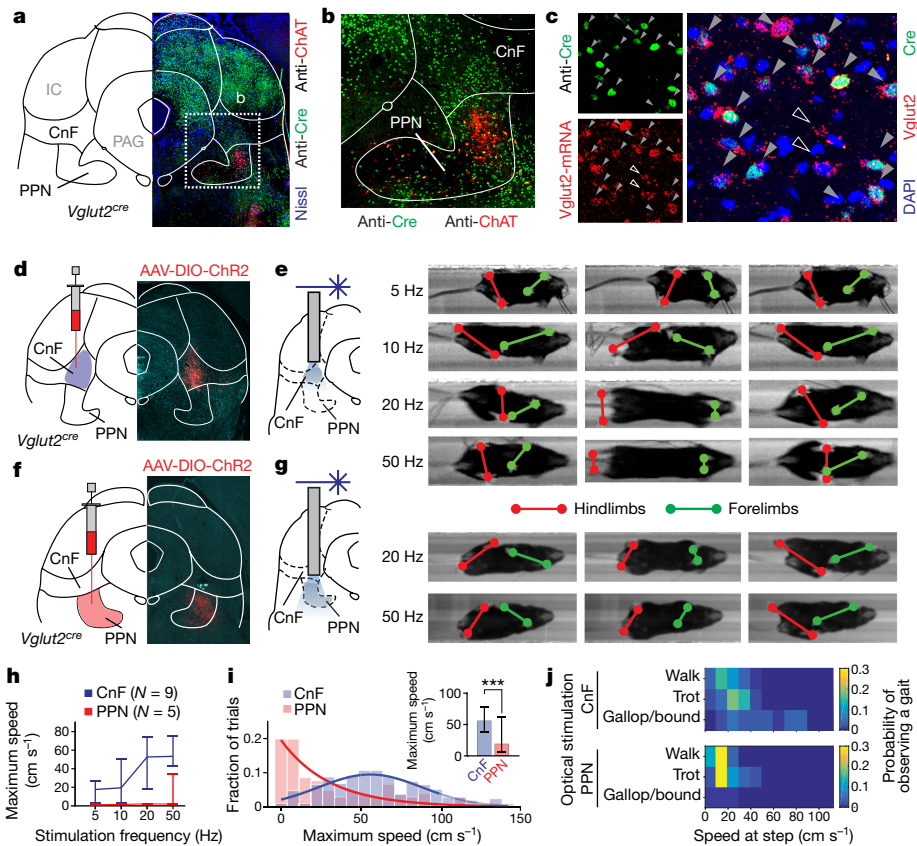


Figure 1 | The control of speed and gait by glutamatergic neurons in the CnF and the PPN. **a–c.** Identification of regions in *Vglut2^{Cre}* mice and localization of neurotransmitters. ChAT, choline acetyltransferase; DAPI, 4',6-diamidino-2-phenylindole; IC, inferior colliculus; PAG, periaqueductal grey. Solid arrowheads in **c** show *Vglut2⁺* cells. **d–g.** Examples of locomotion induced by optical stimulation of the CnF (**d, e**) and the PPN (**f, g**).

CnF neurons produced a wide range of speeds (Fig. 1i, blue) and all gaits: the alternating gaits of walk and trot and the synchronous gaits of gallop and bound (Fig. 1e, j, top)^{1,17,18}. The onset of locomotion was in the range of 100 to 150 ms (Extended Data Fig. 2c, blue line) and remained constant with the variation of stimulation frequency ($P > 0.05$, Kruskal–Wallis test).

Light activation of the *Vglut2⁺*ChR2 PPN neurons also initiated locomotion from rest (Fig. 1f, g, $N = 5$ out of 7 mice; movements detected in 31 out of 67 trials, 46%). Low-frequency stimulation (< 10 Hz) was not able to induce locomotion (Fig. 1h; Extended Data Fig. 2b). Increasing the frequency of stimulation increased the speed of locomotion ($P > 0.05$, Spearman correlation), however very high speeds were not obtained (Fig. 1g, h, red; Extended Data Fig. 2b)—the maximum speed when stimulating *Vglut2⁺*ChR2 PPN neurons was 19 cm s^{-1} , compared with 56 cm s^{-1} for *Vglut2⁺*ChR2 CnF neurons ($P < 0.001$, Mann–Whitney *U*-test; Fig. 1i, Supplementary Video 2). Gallop and bound were also not induced upon increasing the stimulation frequency (Fig. 1g, j, bottom). In addition, the onset of the initiation of locomotion was significantly longer (0.2–1.5 s) after the stimulation of *Vglut2⁺*ChR2 PPN neurons compared with *Vglut2⁺*ChR2 CnF neurons (Extended Data Fig. 2c, red line; $P < 0.05$, Mann–Whitney *U*-test). Stimulation of *Vglut2⁺*ChR2 PPN neurons (expression of ChR2 in Extended Data Fig. 1b) during ongoing locomotion modulated the speed ($P = 0.03$, Wilcoxon signed-rank test, causing an overall increase in speed of 18% compared with that before light onset (Extended Data Fig. 2d). However, the speed after stimulation remained within the ranges of walk and trot, confirming that selective activation of *Vglut2⁺*ChR2 PPN neurons could not initiate fast, synchronous gaits.

h. Maximum speeds evoked by stimulation of the PPN (red; $N = 5$ mice, $n = 67$ trials) and the CnF (blue; $N = 9$ mice, $n = 131$ trials). Error bars indicate the 25th and 75th percentiles of the distribution. **i.** The fraction of trials at a given maximum speed (inset; *** $P < 0.001$, two-tailed Mann–Whitney *U*-test). **j.** Probability of observing different gaits upon optical stimulation of neurons in the CnF (top) or the PPN (bottom).

The frequency of stimulation did not directly translate into the observed stepping frequency. However, the relationship between stepping frequency and the velocity of locomotion (Extended Data Fig. 2e) was similar to that seen during spontaneous locomotion in wild-type mice¹, showing that locomotor activity resulting from light stimulation is similar to that exhibited naturally.

The optogenetically-induced locomotor phenotypes were linked to glutamatergic neurons in PPN or CnF. Locomotion was not induced by stimulation of the local inhibitory neurons in the PPN and the CnF¹⁴, or the cholinergic cells in the PPN, and their activation slowed or stopped ongoing locomotion (Extended Data Fig. 3a–e).

Dual and singular control of locomotion

The optogenetically induced locomotor phenotypes raise the question of whether activity in glutamatergic neurons in both the PPN and the CnF, or in either location independently, is necessary to maintain ongoing locomotion at different speeds. We therefore performed experiments that selectively dampened the activity of the identified populations using the inhibitory muscarinic designer receptor hM4Di (iDREADD), which is activated by clozapine *N*-oxide (CNO)^{19,20}. *Vglut2^{Cre}* mice were bilaterally injected with iDREADDs in both structures (CnF, $N = 8$; PPN, $N = 9$; CnF and PPN, $N = 6$; injection sites shown in Extended Data Fig. 4).

In non-viral-injected mice, no difference was seen in the instantaneous speed attained on a treadmill after treatment with either saline injections or intraperitoneal CNO (1 mg kg^{-1}) (Extended Data Fig. 5a). Test mice with viral infections that received saline attained average speeds of $26\text{--}27 \text{ cm s}^{-1}$ and maximum speeds of $47\text{--}55 \text{ cm s}^{-1}$ (Fig. 2a, b), corresponding to the slow walk/trot and fast trot ranges,

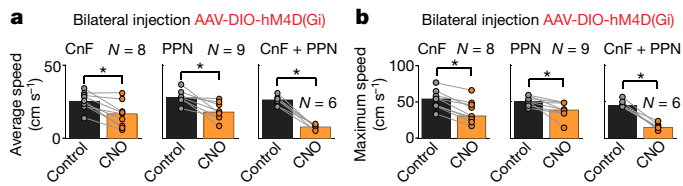


Figure 2 | The PPN and the CnF provide dual control of slower locomotion. **a, b,** Bilateral inhibition of CnF and/or PPN neurons with iDREADDs in *Vglut2^{Cre}* mice. Average (**a**) and maximum (**b**) speed of mice before and after the administration of CNO (* $P < 0.05$, two-tailed Wilcoxon signed-rank test).

respectively, of spontaneous locomotion in untreated adult mice¹. When *Vglut2⁺*iDREADD CnF neurons were inactivated, there was a reduction in both the average and the maximum speed (before versus after CNO treatment, average speed 27 cm s^{-1} versus 20 cm s^{-1} , maximum speed 50 cm s^{-1} versus 41 cm s^{-1} , Mann–Whitney U -test $P < 0.05$, Fig. 2a, b); similar results were obtained when *Vglut2⁺*iDREADD PPN neurons were inhibited (average speed 27 cm s^{-1} versus 18 cm s^{-1} , maximum speed 54 cm s^{-1} versus 43 cm s^{-1} , Mann–Whitney U -test $P < 0.05$, Fig. 2a, b). These effects developed over time, with the maximum effects observed after 30 min (Extended Data Fig. 5b–g). Notably, when the iDREADD virus was injected in both the PPN and the CnF bilaterally, the mice could achieve only very slow forward locomotion—typically single steps with an overall speed within the walking range (Fig. 2a, b).

These experiments suggest that glutamatergic subpopulations in both the PPN and the CnF in conjunction are necessary to maintain ongoing locomotion in the walk and trot range, and that both the PPN and the CnF can independently support slower, alternating locomotion within the walking range.

To investigate the ability of *Vglut2⁺*CnF neurons to initiate the gaits gallop and bound, we first tested the inactivation of *Vglut2⁺*iDREADD CnF neurons in a behavioural assay that allowed for fast, escape-like behaviour (Fig. 3a; Methods). Under control conditions (that is, before CNO treatment), high-speed, escape-like locomotion involving

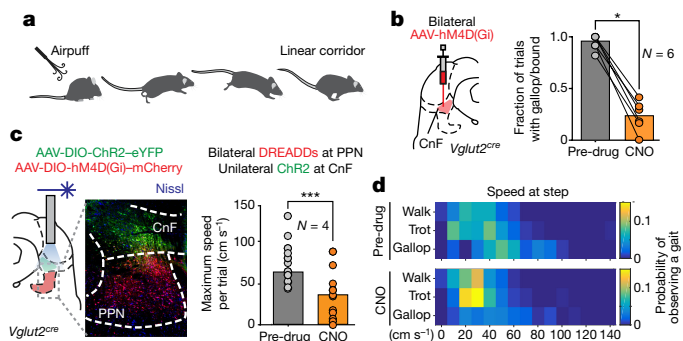


Figure 3 | Glutamatergic neurons in the CnF are required and are sufficient for fast, synchronous locomotion. **a,** To induce fast, escape-like locomotion, air puffs are applied to the back of the mouse when situated at the beginning of a linear corridor. **b,** Bilateral injection of iDREADDs in the CnF of *Vglut2^{Cre}* mice (left) and the probability of evoking the gaits of gallop and/or bound during air-puff-induced escape behaviour, before and after inactivation of the CnF with CNO ($P = 0.0312$, two-tailed Wilcoxon signed-rank test, $N = 6$). **c,** Injection and expression of AAV-ChR2 in the CnF (left, green) and iDREADDs in the PPN (left, red) and fibre placement. Maximum speeds of locomotion (right) combining iDREADDs in the PPN and optogenetic activation of the CnF at 50 Hz, both before (grey; $n = 13$ repetitions, $N = 4$ mice) and after (orange; $n = 12$ repetitions, $N = 4$ mice) the administration of CNO. $P < 0.05$, two-tailed Mann–Whitney U -test. **d,** Probability of observing different gaits upon stimulation of neurons as described in **c**. Drawing in Fig. 3a reproduced with permission from Mattias Karlén.

gallop or bound was observed in 94% of the trials (66 out of 70, $N = 6$, Fig. 3b). After treatment with CNO, the same mice were unable to produce high-speed escape-like actions, and showed no or only isolated signs of gallop or bound in 23% of the trials (18 of 79 trials, $N = 6$, $P < 0.05$, Wilcoxon signed-rank test; Fig. 3b). We next tested if gallop and bound could be initiated upon activation of the *Vglut2⁺* CnF neurons independently of a functioning PPN, by bilateral injection of iDREADDs into *Vglut2⁺*PPN neurons and ChR2 into *Vglut2⁺*CnF neurons (Fig. 3c, $N = 4$). Light activation of *Vglut2⁺*CnF neurons induced a range of locomotor speeds, and all gaits—including gallop and bound—both before and after CNO injection, with only a reduction in the maximum speeds observed after CNO treatment (Fig. 3c, d; Supplementary Video 3). These results show that glutamatergic neurons in the CnF are necessary for producing gallop and bound, and that they can induce these gaits independently of the glutamatergic neurons in the PPN.

Neuronal firing and its relationship to speed

The complementary roles of glutamatergic neurons in the CnF and the PPN in regulating the speed of alternating locomotion may be reflected in their firing activity. We therefore recorded the activity of CnF and PPN neurons extracellularly when mice were walking or trotting on a treadmill (0 – 30 cm s^{-1}). Glutamatergic neurons were infected with AAV-DIO-ChR2 in either the CnF ($N = 2$) or the PPN ($N = 2$), and identified as infected by their short latency (up to 5 ms) and constant jitter responses to brief pulses of blue light (Fig. 4; Extended Data Fig. 6a). We recorded from a total of 169 *Vglut2⁺*CnF neurons and 493 *Vglut2⁺*PPN neurons; Figure 4a, b shows example neurons in the two structures. The *Vglut2⁺*ChR2 CnF neuron (Fig. 4a) showed a notable correlation between speed and firing rate. The depicted *Vglut2⁺*ChR2 PPN neurons were recruited at the beginning of the locomotor bout

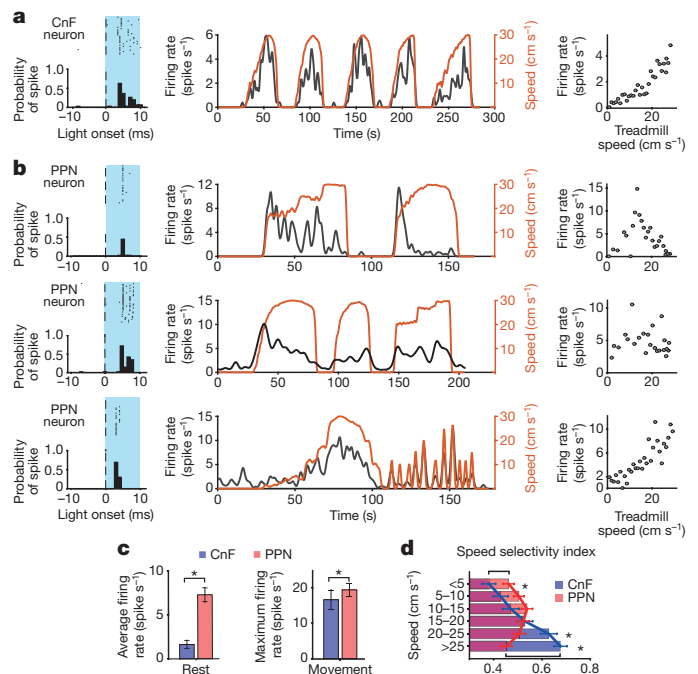


Figure 4 | Coding of speed in glutamatergic neurons in the CnF and the PPN. **a, b,** Firing of example neurons at different locomotion speeds, recorded in the CnF (**a**) and the PPN (**b**). **c,** Average (left, at rest) and maximum (right, during movement) neuronal firing rates in the CnF ($n = 79$) and the PPN ($n = 105$). $P < 0.001$, two-tailed Mann–Whitney U -test. **d,** Speed selectivity index. *Vglut2⁺*PPN neurons were more selective at lower speeds, whereas *Vglut2⁺*CnF neurons were more selective at higher speeds. * $P < 0.05$, two-tailed Mann–Whitney U -test with post hoc Bonferroni correction. Data are mean \pm s.e.m.

and then slowly derecruited (Fig. 4b, top), showed no modulation with speed (Fig. 4b, middle), or showed a clear modulation with the speed of locomotion (Fig. 4b, bottom).

For further quantitative analysis, we considered only the glutamatergic neurons in the PPN and the CnF for which the firing rate was modified upon changes in speed (Spearman correlation $P < 0.01$; PPN, $n = 105$, median correlation 0.63; CnF, $n = 79$, median correlation 0.63) (Extended Data Fig. 6b). Among these cells, differences were observed between the firing distributions of Vglut2⁺ChR2 CnF neurons and Vglut2⁺ChR2 PPN neurons during rest and movement (Fig. 4c, rest: average activity CnF 1.62 versus PPN 7.27; movement: maximum activity CnF 16.75 versus PPN 19.53; both $P < 0.05$, Mann–Whitney U -test).

We quantified these firing profiles by computing a speed selectivity index, which weights how much stronger the firing rate is at a specific speed compared to the activity at rest (Fig. 4d). Neurons in both the CnF and the PPN showed selectivity with respect to their baselines (Fig. 4d, $P < 0.05$, Wilcoxon signed-rank test against baseline with post hoc Bonferroni correction). Nevertheless, the selectivity was different: Vglut2⁺ChR2 PPN neurons were more selective at the lowest treadmill speed (below 5 cm s⁻¹) whereas Vglut2⁺ChR2 CnF neurons were more selective at the highest treadmill speed (above 20 cm s⁻¹) ($P < 0.05$, Mann–Whitney U -test with post hoc Bonferroni correction).

The relationship between firing rate and speed supports the suggestion that glutamatergic neurons in both the CnF and the PPN contribute towards programming the speed of alternating gait locomotion. At the lowest speeds, the PPN neurons have a greater contribution than those of the CnF, whereas the CnF neurons show the strongest contribution at higher speeds.

The PPN is involved in exploratory behaviour

The different firing behaviour of the PPN and the CnF neurons raises the possibility that they might be mobilized differently to support slow, explorative behaviour. We therefore measured explorative behaviour using the hole-board test^{21,22} (Fig. 5a), a context that encourages slow-speed locomotion for exploratory purposes. Mice were injected bilaterally with iDREADDs targeting Vglut2⁺ neurons in either the CnF or the PPN (Fig. 5b, c). Changes in locomotion induced by Vglut2⁺iDREADD CnF neurons or Vglut2⁺iDREADD PPN neurons were measured by the average speed of locomotion, the distance travelled and the ambulation time in the same mouse after the injection of either saline or CNO. CnF-injected mice ($N = 6$) did not show any differences in these locomotor parameters (Wilcoxon signed-rank test, saline versus CNO, $P > 0.05$), whereas PPN-injected mice ($N = 6$) showed a significant reduction in the total distance travelled and the average speed (Wilcoxon signed-rank test, saline versus CNO, $P < 0.05$) (data not shown). As a measure of exploration, we measured the number and the fraction of time of head-dips. Before and after the inactivation of Vglut2⁺iDREADD CnF neurons, there was no difference in these parameters (Fig. 5b, $N = 6$; $P > 0.05$, Wilcoxon signed-rank test), however both were significantly reduced upon the inactivation of Vglut2⁺iDREADD PPN neurons, (Fig. 5c, $N = 6$; $P < 0.05$, Wilcoxon signed-rank test). These results support the suggestion that glutamatergic PPN activity may facilitate slow, explorative locomotor behaviour.

Next, we tested whether PPN activation could also increase exploration. Vglut2⁺ neurons in the CnF or the PPN ($N = 2$ and 4, respectively; Extended Data Fig. 7) were infected with ChR2 (Fig. 5d, e) and stimulated for 10 s (40 Hz) at random times throughout the five-minute exploration period (Supplementary Video 4). There was a significant reduction in head-dipping before and after stimulation of the CnF (Fig. 5d, $P < 0.05$, Mann–Whitney U -test, $n = 40$ repetitions in $N = 2$ mice)—due to the induction of escape-like behaviour—but a significant increase in both the number and the fraction of time of head-dips during stimulation of the PPN (Fig. 5e, $P < 0.05$, Mann–Whitney U -test, $n = 53$ repetitions in $N = 4$ mice). These experiments further support the idea that activity in Vglut2⁺ PPN neurons facilitates movements at slow speeds for the purpose of explorative behaviour.

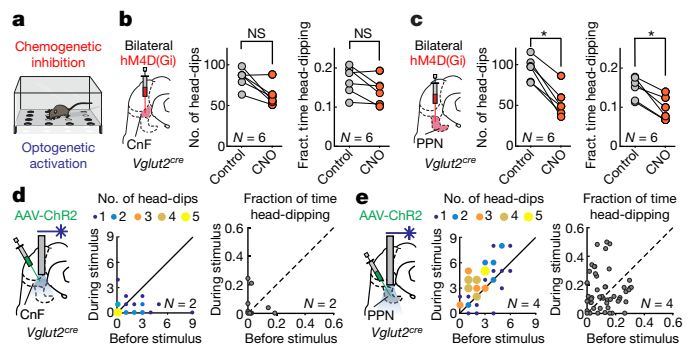


Figure 5 | Selection of exploration from the PPN. **a**, Set-up of the exploratory hole-board experiments. **b**, Bilateral inactivation of the CnF (left, $N = 6$) did not reduce either the frequency of head-dips (middle, $P > 0.05$) or the fraction of time spent head-dipping (right, $P > 0.05$). **c**, Bilateral inactivation of the PPN (left, $N = 6$) reduced both the frequency of head-dips (middle, $P = 0.031$) and the fraction of time spent head-dipping (right, $P = 0.031$). **d**, Optogenetic stimulation of the CnF (left, $N = 2$) induced a decrease in the number of head-dips (middle, $P = 0.0023$) but not in the fraction of time spent head-dipping (right, $P > 0.05$). **e**, Stimulation of the PPN (left, $N = 4$) increased both the number of head-dips (middle, $P < 0.001$) and the fraction of time spent head-dipping (right, $P = 0.0218$). All statistical tests were two-tailed Wilcoxon signed-rank tests. Drawing in Fig. 5a reproduced with permission from Mattias Karlén.

Brain-wide inputs to the CnF and the PPN

To investigate the regulation of glutamatergic excitatory neurons of the CnF and the PPN, we traced the sources of neuronal inputs into each structure using rabies-based mono-synaptically restricted retrograde trans-synaptic circuit tracing (refs 23, 24; Methods; Fig. 6). Trans-synaptically labelled neurons are visualized as red-only neurons in Fig. 6a. The overall distribution of projecting neurons to the Vglut2⁺CnF neurons or the Vglut2⁺PPN neurons was visibly different (orange dots in Fig. 6b; PPN, $N = 3$; CnF, $N = 3$). Most inputs were ipsilateral to the injection site, and inputs to Vglut2⁺CnF neurons were more restricted compared to those of the Vglut2⁺PPN neurons. The main inputs to Vglut2⁺PPN neurons originate in midbrain structures (Fig. 6c) and sensory-motor and raphe nuclei in the brainstem (Fig. 6d). Furthermore, Vglut2⁺PPN neurons also receive direct input from the output nuclei in the basal ganglia (Fig. 6e, f). Sparse inputs were found from sensory-motor and frontal cortices or the hypothalamus (Fig. 6c). Therefore, Vglut2⁺PPN neurons integrate sensory-motor information from many brain structures. Conversely, Vglut2⁺CnF neurons receive little input from basal ganglia output nuclei (Fig. 6e, f) or from cortices, but stronger projections from midbrain structures (for example the periaqueductal grey or the inferior colliculus, Fig. 6c, d) that have been assigned a role in escape responses^{25,26}.

Lastly, Vglut2⁺ neurons in the CnF and the PPN have reciprocal projections, with dominant projections from the CnF to the PPN (Extended Data Fig. 8); these provide gateways for Vglut2⁺CnF neurons to modulate PPN neurons in the range of slower, alternating locomotion.

Convergent and divergent outputs

Descending projections from Vglut2⁺CnF neurons and Vglut2⁺PPN neurons were evaluated using transmitter-specific anterograde tracing (Extended Data Fig. 9a). Few neurons projected directly to the cervical and thoracic spinal cord (see also refs 27–29) (Extended Data Fig. 9c–5). Vglut2⁺PPN neurons have broad—predominantly ipsilateral—projections, including to motor-related nuclei in the pons as well as to modulatory nuclei (Extended Data Fig. 9b, c1–4). Most of these brainstem nuclei project to the spinal cord in mice²⁷. By contrast, the CnF has more restricted projection, and both overlapping and non-overlapping projections with the PPN in the medulla (Extended Data Fig. 9c1–4).

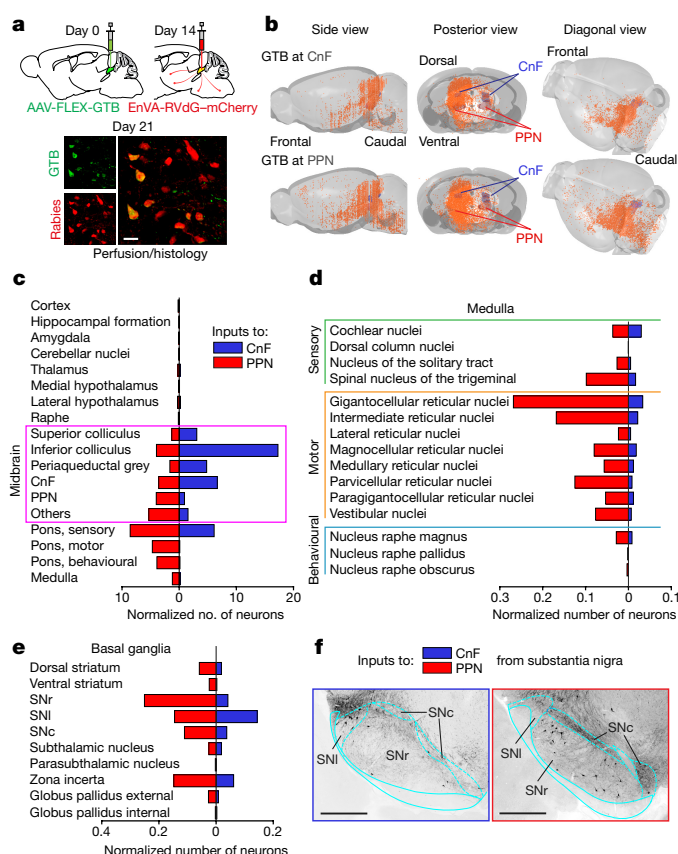


Figure 6 | Neurons in the CnF and the PPN have differential input matrices. **a**, Timeline of mono-synaptically restricted trans-synaptic retrograde tracing. Bottom, neurons infected with GTB (green, upper left), or rabies virus (red, lower left). Neurons co-infected with GTB and rabies shown in yellow. **b**, Reconstruction of projections to the CnF (top) and the PPN (bottom), as revealed by monosynaptically restricted trans-synaptic retrograde labelling ($N = 3$). **c–e**, Regional distribution (median, $N = 3$) of neurons projecting to the CnF (blue) and to the PPN (red), normalized to the number of primary infected neurons in either structure. **f**, Examples of labelled neurons (black) in the substantia nigra projecting onto glutamatergic CnF (left) or PPN (right) neurons. Scale bars: **a**, 20 μm ; **f**, 500 μm . SNc, substantia nigra pars compacta; SNI, substantia nigra pars lateralis; SNr, substantia nigra pars reticulata. The mouse brain schematics in this figure have been reproduced with permission from Elsevier³⁸.

Conclusions

Our study shows that two transmitter-defined and spatially segregated populations of neurons in the mouse midbrain form command pathways that encode speeds of locomotion in complementary ways. Neuronal circuits in the PPN and the CnF both contribute to the maintenance and speed regulation of slower locomotion, whereas only the CnF is able to elicit high-speed, synchronous locomotor activity. The functional locomotor signatures are linked to the activity of the glutamatergic neurons in the CnF and the PPN. The focus on speed control and the selection of gaits provide a combined solution to understanding the functional organization of the midbrain structures involved in locomotor control. The concept of a unitary mesencephalic locomotor region in mammals is therefore refined by a more advanced model, in which the locomotor control function resides in both the PPN and the CnF.

The support of slow explorative and fast escape behaviour by glutamatergic PPN and CnF neurons, respectively, suggests that these neuronal circuits may be recruited in specific behavioural contexts. The differential input matrices into glutamatergic neurons in the CnF or the PPN also suggest the existence of dual functions in addition to the combined control of alternating gaits. The strong inputs into

Vglut2⁺CnF neurons from the periaqueductal grey (especially the dorsal part), the inferior colliculus and the hypothalamus are in accordance with previous anatomical findings¹² and suggest that CnF-mediated fast locomotion may be generated as part of an escape response independent of the PPN. As previously shown, PPN neurons receive rich projections from basal ganglia nuclei^{12,14}, but also from many midbrain and medullary sensory-motor nuclei as well as from the motor cortex. This innervation pattern is in accordance with a role of glutamatergic PPN neurons in exploratory locomotor behaviour under the motor action selection of the basal ganglia^{7,30–34}. The strong connection from the basal ganglia also suggests that dysregulation of glutamatergic neurons in the PPN may have important roles in locomotor disability related to Parkinson's disease.

The descending projections from glutamatergic CnF and PPN neurons suggest that the speed signal is funnelled through diverse brainstem nuclei, which in turn project to the locomotor networks in the spinal cord. The convergent projections of the CnF and the PPN to regions that contain excitatory reticulospinal neurons^{35,36} provide a gateway to support alternating gaits in a speed-dependent manner³⁷. This area may also be accessed from the CnF independent of the PPN, as the CnF can initiate gallop and bound without activity in the PPN. Conversely, neurons in the PPN project more broadly to nuclei in the pons and the medulla, which are mostly devoid of CnF projections, and may provide descending pathway(s) involved in slow, explorative locomotor behaviour.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 February; accepted 8 December 2017.

Published online 17 January 2018.

- Bellardita, C. & Kiehn, O. Phenotypic characterization of speed-associated gait changes in mice reveals modular organization of locomotor networks. *Curr. Biol.* **25**, 1426–1436 (2015).
- Grillner, S. The motor infrastructure: from ion channels to neuronal networks. *Nat. Rev. Neurosci.* **4**, 573–586 (2003).
- Kiehn, O. Decoding the organization of spinal circuits that control locomotion. *Nat. Rev. Neurosci.* **17**, 224–238 (2016).
- Grillner, S. & Jessell, T. M. Measured motion: searching for simplicity in spinal locomotor networks. *Curr. Opin. Neurobiol.* **19**, 572–586 (2009).
- Brownstone, R. M. & Wilson, J. M. Strategies for delineating spinal locomotor rhythm-generating networks and the possible role of Hb9 interneurons in rhythmogenesis. *Brain Res. Rev.* **57**, 64–76 (2008).
- Goulding, M. Circuits controlling vertebrate locomotion: moving in a new direction. *Nat. Rev. Neurosci.* **10**, 507–518 (2009).
- Jordan, L. M., Liu, J., Hedlund, P. B., Akay, T. & Pearson, K. G. Descending command systems for the initiation of locomotion in mammals. *Brain Res. Rev.* **57**, 183–191 (2008).
- Dubuc, R. *et al.* Initiation of locomotion in lampreys. *Brain Res. Rev.* **57**, 172–182 (2008).
- Takakusaki, K., Chiba, R., Nozu, T. & Okumura, T. Brainstem control of locomotion and muscle tone with special reference to the role of the mesopontine tegmentum and medullary reticulospinal systems. *J. Neural Transm. (Vienna)* **123**, 695–729 (2016).
- Shik, M. L., Severin, F. V. & Orlovsky, G. N. Control of walking and running by means of electrical stimulation of the mesencephalon. *Electroencephalogr. Clin. Neurophysiol.* **26**, 549 (1969).
- Garcia-Rill, E., Hyde, J., Kezunovic, N., Urbano, F. J. & Petersen, E. The physiology of the pedunculopontine nucleus: implications for deep brain stimulation. *J. Neural Transm. (Vienna)* **122**, 225–235 (2015).
- Ryczko, D. & Dubuc, R. The multifunctional mesencephalic locomotor region. *Curr. Pharm. Des.* **19**, 4448–4470 (2013).
- Martinez-Gonzalez, C., Bolam, J. P. & Mena-Segovia, J. Topographical organization of the pedunculopontine nucleus. *Front. Neuroanat.* **5**, 22 (2011).
- Roseberry, T. K. *et al.* Cell-type-specific control of brainstem locomotor circuits by basal ganglia. *Cell* **164**, 526–537 (2016).
- Wang, H. L. & Morales, M. Pedunculopontine and laterodorsal tegmental nuclei contain distinct populations of cholinergic, glutamatergic and GABAergic neurons in the rat. *Eur. J. Neurosci.* **29**, 340–358 (2009).
- Borgius, L., Restrepo, C. E., Leao, R. N., Saleh, N. & Kiehn, O. A transgenic mouse line for molecular genetic analysis of excitatory glutamatergic neurons. *Mol. Cell. Neurosci.* **45**, 245–257 (2010).
- Machado, A. S., Darmohray, D. M., Fayad, J., Marques, H. G. & Carey, M. R. A quantitative framework for whole-body coordination reveals specific deficits in freely walking ataxic mice. *eLife* **4**, e07892 (2015).

18. Lemieux, M., Josset, N., Roussel, M., Couraud, S. & Bretzner, F. Speed-dependent modulation of the locomotor behavior in adult mice reveals attractor and transitional gaits. *Front. Neurosci.* **10**, 42 (2016).
19. Roth, B. L. DREADDs for neuroscientists. *Neuron* **89**, 683–694 (2016).
20. Sternson, S. M. & Roth, B. L. Chemogenetic tools to interrogate brain functions. *Annu. Rev. Neurosci.* **37**, 387–407 (2014).
21. Kliethermes, C. L. & Crabbe, J. C. Pharmacological and genetic influences on hole-board behaviors in mice. *Pharmacol. Biochem. Behav.* **85**, 57–65 (2006).
22. File, S. E. & Wardill, A. G. Validity of head-dipping as a measure of exploration in a modified hole-board. *Psychopharmacologia* **44**, 53–59 (1975).
23. Callaway, E. M. & Luo, L. Monosynaptic circuit tracing with glycoprotein-deleted rabies viruses. *J. Neurosci.* **35**, 8979–8985 (2015).
24. Wall, N. R., Wickersham, I. R., Cetin, A., De La Parra, M. & Callaway, E. M. Monosynaptic circuit tracing *in vivo* through Cre-dependent targeting and complementation of modified rabies virus. *Proc. Natl Acad. Sci. USA* **107**, 21848–21853 (2010).
25. Brandão, M. L., Anseloni, V. Z., Pandóssio, J. E., De Araújo, J. E. & Castilho, V. M. Neurochemical mechanisms of the defensive behavior in the dorsal midbrain. *Neurosci. Biobehav. Rev.* **23**, 863–875 (1999).
26. Lovick, T. A. The periaqueductal gray-rostral medulla connection in the defence reaction: efferent pathways and descending control mechanisms. *Behav. Brain Res.* **58**, 19–25 (1993).
27. Liang, H., Paxinos, G. & Watson, C. Spinal projections from the presumptive midbrain locomotor region in the mouse. *Brain Struct. Funct.* **217**, 211–219 (2012).
28. Ryczko, D. *et al.* Forebrain dopamine neurons project down to a brainstem region controlling locomotion. *Proc. Natl Acad. Sci. USA* **110**, E3235–E3242 (2013).
29. Skinner, R. D., Kinjo, N., Ishikawa, Y., Biedermann, J. A. & Garcia-Rill, E. Locomotor projections from the pedunculopontine nucleus to the medioventral medulla. *Neuroreport* **1**, 207–210 (1990).
30. Costa, R. M. Plastic corticostriatal circuits for action learning: what's dopamine got to do with it? *Ann. N. Y. Acad. Sci.* **1104**, 172–191 (2007).
31. Grillner, S., Robertson, B. & Stephenson-Jones, M. The evolutionary origin of the vertebrate basal ganglia and its role in action selection. *J. Physiol. (Lond.)* **591**, 5425–5431 (2013).
32. Friend, D. M. & Kravitz, A. V. Working together: basal ganglia pathways in action selection. *Trends Neurosci.* **37**, 301–303 (2014).
33. Sinnamoni, H. M. Preoptic and hypothalamic neurons and the initiation of locomotion in the anesthetized rat. *Prog. Neurobiol.* **41**, 323–344 (1993).
34. Graybiel, A. M. & Grafton, S. T. The striatum: where skills and habits meet. *Cold Spring Harb. Perspect. Biol.* **7**, a021691 (2015).
35. Noga, B. R., Kriellaars, D. J., Brownstone, R. M. & Jordan, L. M. Mechanism for activation of locomotor centers in the spinal cord by stimulation of the mesencephalic locomotor region. *J. Neurophysiol.* **90**, 1464–1478 (2003).
36. Drew, T., Dubuc, R. & Rossignol, S. Discharge patterns of reticulospinal and other reticular neurons in chronic, unrestrained cats walking on a treadmill. *J. Neurophysiol.* **55**, 375–401 (1986).
37. Capelli, P., Pivetta, C., Esposito, M. S. & Arber, S. Locomotor speed control circuits in the caudal brainstem. *Nature* **551**, 373–377 (2017).
38. Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 4th edn (Academic, 2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported by European Research Council grant ERC-693038 (O.K.), NINDs NS 090919 (O.K.), The Swedish Medical Research Council (O.K., G.F.), StratNeuro (O.K.) and Novo Nordisk Foundation Laureate Research grant NNF 15OC0014186 (O.K.). We thank P. Löw for assistance with viral work and K. Deisseroth for providing viral ChR2 vectors.

Author Contributions O.K. initiated the project. V.Cag., R.L., H.G.-E. and O.K. designed the experiments with contributions from all authors. V.Cag. and R.L. performed optogenetic experiments, *in vivo* recordings and analysis. C.B. and H.G.-E. contributed to locomotor gait analysis, and J.B. to the initial optogenetic experiments. H.G.-E. and D.M. were responsible for chemogenetic inactivation experiments together with R.L. and V.Cag., and all analysed the data together with O.K. H.G.-E. and R.L. performed anatomical analysis with V.Cag. V.Cal. carried out *in situ* hybridizations. V.Cag. and O.K. wrote the paper with contributions from all authors. O.K. supervised all aspects of the work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to O.K. (Ole.Kiehn@ki.se), V.Cag (caggiano@gmail.com) or R.L. (Roberto.Leiras@sund.ku.dk).

METHODS

Data reporting. The experiments were not randomized. For the hole-board experiments, the investigators were blinded to treatment allocation and outcome assessment. For all other experiments, the investigators were not blinded to allocation during experiments and outcome assessment. No statistical methods were used to predetermine sample size.

Mice. All experiments were approved by the local ethical committee (Stockholm Norra Djuretiska nämnden). For most experiments, adult *Vglut2^{cre}* transgenic mice¹⁶ were used (3–5 months old, of both sexes). In some experiments, adult *Vgat^{cre}* and *Chat^{cre}* (ChAT-IRES-Cre knock-in, Jackson Laboratory) transgenic mice were used (8–14 weeks old, of both sexes). *Chat^{cre}* mice were crossed with Rosa26-CAG-LSL-ChR2-eYFP-WPRE mice (Jackson Laboratory). Mice were genotyped before the experiments.

In vivo optogenetic experiments. For viral transfection of *Vglut2*-expressing neurons, *Vglut2^{cre}* mice aged 3–5 months were anaesthetized with isoflurane. For activation experiments, 100–300 nl of an AAVdj-EF1a-DIO-hChR2-p2A-mCherry-WPRE virus was pressure-injected using a glass micropipette into the CnF (anteroposterior angle 15°, from bregma: anteroposterior –5.7 mm, mediolateral 1.2 mm, depth 2.9 mm) or the PPN (anteroposterior angle 20°, from bregma: anteroposterior –5.9 mm, mediolateral 1.2 mm, depth 4.2 mm). In the same surgery, an optical fibre (200 µm core, numerical aperture 0.22, Thorlabs) held in a 1.25 mm ferrule was implanted (500 µm above the injection site) for stimulation of the transfected cells. To reduce firing in *Vglut2*-expressing neurons, 100–200 nl of an AAV-hSyn-DIO-hM4D(Gi)-mCherry virus (UNC vector core) was bilaterally injected in either the CnF or the PPN, or in both structures.

When accessing the PPN, great care was taken not to damage the CnF by adjusting the angle to 20°. By measuring the response evoked from stimulation of the CnF in mice ($N=2$) expressing ChR2 in both the CnF and the PPN, we confirmed that acutely lowering the optical fibre to stimulate first the CnF and then the PPN did not damage the CnF. Thus, the same activation of both the CnF and the PPN was obtained both when lowering and retracting the probe, demonstrating that damage to the CnF did not account for the findings in the PPN.

Some mice were injected bilaterally with AAV-hSyn-DIO-hM4D(Gi)-mCherry virus in the PPN, and unilaterally with AAVdj-EF1a-DIO-hChR2-p2A-eYFP-WPRE and implanted for optical stimulation of the CnF. For the first week after surgery, all mice were treated daily with analgesics and monitored for any sign of discomfort.

Optogenetic stimulation. A 473 nm laser (Optoduet, Ikecool Corporation) was connected to the ferrule that was chronically implanted on the mice through a ceramic mating sleeve. For light-activation of ChR2-transfected neurons, we used trains of light pulses (Master-8 pulse generator, AMPI or custom-made MATLAB (Mathworks Inc.) scripts) with variable pulse durations and frequencies. When the frequency was changed, the pulse duration was also changed to obtain the same intensity of stimulation with constant laser power. The intensity of the laser was between 5 mW and 30 mW.

Drugs. Clozapine *N*-oxide (CNO, Sigma-Aldrich) was dissolved in physiological saline to obtain a final dose of 1 mg kg⁻¹, before intraperitoneal injection.

Behavioural test in a corridor. Locomotor behaviour was recorded with the TSE MotoRater system with the mice running spontaneously on a 1.2-m long runway, as previously described^{1,39}. Videos were acquired using a high-speed camera at 300 frames per second, and analysed offline.

For the induction of fast, escape-like locomotion (gallop and bound), we used standardized air puffs (50 psi, 500 ms long) applied to the back of the mouse when it was situated at the beginning of the corridor. The test was repeated ten times with several minutes of rest between trials, both before and after intraperitoneal injection of CNO.

Behavioural test on a treadmill. Locomotion was analysed using a motorized transparent treadmill with adjustable speed range (Exer Gait XL, Columbus Instruments). The mice were conditioned to locomote on the treadmill set at constant speed, in bouts of 20 s separated by 1–2 min inter-trial periods. Ventral plane videography was recorded at 100 frames per second. Each mouse was tested at three different speeds: 0–4 cm s⁻¹, 4–20 cm s⁻¹, and >20 cm s⁻¹ before and after intraperitoneal injection of CNO. The instantaneous speed of the mice was measured throughout the experiments using custom-made MATLAB scripts with foot placement monitored from below.

Hole-board behavioural test. Exploratory behaviour was analysed using a modified version of the hole-board apparatus, consisting of test boxes made of transparent Plexiglas (45 cm × 45 cm × 41 cm) and a hole-board frame with 16 holes in a grid-pattern (2 cm diameter, 9 cm apart), placed 4 cm above the floor of the testing box. The apparatus was located in a testing room with dimmed illumination (40 lux). Odour-impregnated bedding from cages of the same gender, which is a strong exploratory motivator, was placed below the hole-board frame. To reduce habituation due to multiple trials, new social odour sources were placed

under the hole-board platform for every new trial. During the experiments, the experimenters knew whether they had injected saline or CNO, but were blind to whether the mouse had the virus injected or not; the experimenter also did not know the site of injection (that is, either CnF or PPN).

The same hole-board set-up was used to induce exploration with the stimulation of channelrhodopsin-expressing *Vglut2⁺* neurons. Mice were first tested with the MotoRater and light-activated responders were pre-selected for exploration tests. On the test day, they were placed in the open field and stimulus parameters were adjusted for each mouse in order to produce a locomotor response (typically 30–40 Hz, pulse duration 10 ms). The mice were then stimulated with trains of stimuli lasting 10 s, delivered at random intervals every 20–40 s over the five-minute test period.

Monosynaptically restricted trans-synaptic labelling. We used a glycoprotein (G)-deleted rabies virus^{23,24} pseudotyped with the envelope glycoprotein EnvA to enable the selective infection of glutamatergic cells via the TVA receptor. The TVA receptor was delivered together with the rabies glycoprotein conditionally to glutamatergic cells, by injecting 200–300 nl AAVdj-EF1a-FLEX-GTB virus (helper virus, Salk Institute, visualized in green in Fig. 6a) into either the CnF or the PPN in *Vglut2^{cre}* mice. Two weeks after the helper virus injection, 200–300 nl of an EnvA G-deleted rabies-mCherry conjugate (Salk Institute) was injected at the same location. Finally, one week after the injection of the rabies virus, mice were transcardially perfused and the tissue analysed (see 'Sectioning, histology and imaging').

Anterograde labelling. For anterograde labelling, 50–100 nl of cell-filling AAVdj-EF1a-DIO-hChR2-p2A-mCherry-WPRE and AAVdj-EF1a-DIO-hChR2-p2A-eYFP were injected into the CnF and the PPN, respectively. The mice were euthanized six weeks after the injection.

Sectioning, histology, and imaging. Adult mice were anaesthetized with pentobarbital and perfused with 4% (w/v) paraformaldehyde in PBS. Brains and spinal cords were removed and post-fixed for 3 h in 4% paraformaldehyde. After fixation, tissues were rinsed in PBS, cryoprotected in 25% (w/v) sucrose in PBS overnight and frozen in Neg-50 embedding medium. Coronal sections (30–40 µm thick) were cut on a cryostat.

Sections were permeabilized with PBS and 0.5% (w/v) Triton X-100 (PBST) and blocked in PBST supplemented with 5% (v/v) normal donkey serum (Jackson ImmunoResearch), before incubation for 24–48 h at 4 °C with one or several of the following primary antibodies diluted in PBST supplemented with 1% normal donkey serum: chicken anti-GFP (1:1,000, Abcam, ab13970), rabbit anti-mCherry (1:1,000, Clontech 632496), goat anti-ChAT (1:100, Millipore AB144P), rabbit anti-Cre (1:8,000, a gift from G. Shutz—see ref. 16). Secondary antibodies (F(ab')₂ fragments) were obtained from Jackson ImmunoResearch or Invitrogen, used at 1:500 and incubated for 3 h at room temperature in PBST 1% normal donkey serum. A fluorescent Nissl stain (NeuroTrace Blue 435/455, 1:200, Life Technologies) was added during the primary antibody incubation. No antibody was required to detect the rabies-mCherry labelling. Slides were rinsed, mounted in Prolong Diamond Antifade mounting medium (Life Technologies) and scanned on a confocal laser scanning microscope (LSM510 or LMS700, Zeiss Microsystems) using 10×, 20× and 40× objectives.

Fluorescent *in situ* hybridization combined with immunofluorescence labelling was performed as previously described¹⁶ using a *Vglut2* probe spanning the base pairs 540–983 (produced by L. Borgius).

Assessment of fibre placement and viral expression pattern. The assessment of the position of the optical fibre tip was based on the visible tract in the tissue. The extent of virus expression in *Vglut2^{cre}* or *Vgat^{cre}* mice was evaluated by outlining the area of expression on sections from individual mice redrawn from a mouse brain atlas, and then superimposing all mice at 30% transparency to highlight the average expression in each group (see ref. 40). Mice with no successful bilateral injections in the DREADD experiments were excluded from the analysis.

Trans-synaptic labelling experiments. For trans-synaptic labelling experiments, all sections were serially collected spanning the whole brain, from the C1 vertebral level to the olfactory bulbs. Every third section was scanned for analysis. Each slice was captured with at least two channels: one for the Nissl staining, and the other for the mCherry that enables the detection of rabies-infected neurons. In addition, a third channel was used to detect the GTB in primary-infected neurons at the site of injection. The analysis consisted of two parts. First, anatomical landmarks were identified based on the Nissl staining and matched (affine transformation followed by cubic B-spline transformation) to the coordinate framework (CCF v3) of the Allen Mouse Brain Atlas at 25-µm resolution with custom-made MATLAB scripts. Second, single neurons were automatically detected based on pixel values above the first of eight thresholds computed using Otsu's method. Then, the sections were manually checked to remove fluorescent counts that were inaccurately detected as neurons or to add neurons that were not detected automatically. Projection to the standardized Allen Mouse Brain Atlas was performed via the B-spline maps

computed in the first step. A contrast enhancement and a noise reduction filter were applied using ImageJ to images for publication.

Gait data analysis. Videos were analysed using scripts written in MATLAB. The speed of the mice was detected by colour segmentation with respect to the background and compensated for the movement of the camera in the corridor using the Lucas–Kanade method. The initiation of locomotion was defined as mouse displacement with speeds greater than 3 cm s^{-1} . Gait analysis was performed with the same methods as described previously^{1,39}. A step cycle was defined as a complete cycle of leg movement from the beginning of the stance phase (foot touchdown) to the end of the swing phase (foot touchdown again). The step frequency was defined as the inverse of the step-cycle duration. All steps were divided into four main gaits on the basis of footprint analysis. The classification of steps involved visual inspection followed by quantitative evaluation of limb coordination. For quantification, we identified the beginning of the stance phase (touchdown of the foot with the ground) and the beginning of the swing phase (lift-off of the foot from the ground) for all limbs in each step. Walk is defined as a pattern of limb movement in which three or four feet are on the ground simultaneously (speed $< 25\text{--}30 \text{ cm s}^{-1}$) (ref. 1). Trot is characterized by a pattern of movement in which diagonal pairs of limbs (for example, left forelimb and right hindlimb) move forward simultaneously and homologous pairs of limbs (for example, hindlimbs) are in alternation (speed $30\text{--}70 \text{ cm s}^{-1}$) (ref. 1). Bound is a pattern of movement in which the mouse moves the forelimbs and hindlimbs in synchrony throughout the movement, but with the fore- and hindlimb moving out of phase (speed $80\text{--}150 \text{ cm s}^{-1}$) (ref. 1). Gallop is characterized by synchronized hindlimb movement and out-of-phase forelimb movement (speed $60\text{--}120 \text{ cm s}^{-1}$) (ref. 1).

Neuronal recordings and analysis. Linear arrays (NeuroNexus multi-site electrode, A1-X16-5 mm-100-413) were inserted into the CnF or the PPN through a microscope. Mice were placed on a custom-built treadmill, the speed of which could be continuously changed. Movement of the treadmill, laser stimulation and array data were stored at 25 kHz on a TDT logger and analysed offline. The maximum speed of the treadmill that mice could reliably follow in a head-fixed experimental set-up was 30 cm s^{-1} . Spike sorting was performed offline by adjusting the energy level in a superparamagnetic clustering algorithm (wave_clus⁴¹, https://github.com/csn-le/wave_clus). Spike trains were aligned either to the speed of the treadmill or to the onset of the optical stimulation. Neurons infected with ChR2 were detected by their fast and reproducible response to 20 ms pulses of blue light. The neuronal activity was quantified in a window from 10 ms before light onset to 5 ms after light onset. Neurons that showed a significant increase in the instantaneous frequency of firing in the ‘after-light-onset-period’ compared to the ‘before-light-onset-period’ ($P < 0.05$, Wilcoxon signed-rank test) and had a short-latency response were considered Vglut2⁺ChR2 CnF or Vglut2⁺ChR2

PPN neurons. We calculated the instantaneous frequency of firing and speed of locomotion in 500 ms bins and quantified the relationship between the firing rate and the speed of the treadmill by averaging the firing rate every 1 cm s^{-1} . A neuron was included as speed-related when it showed a significant correlation between the firing rate and the speed of the treadmill ($P < 0.01$, Spearman correlation). A speed selectivity index was calculated as the absolute value of the average binned neuronal activity in specific speed ranges (for example, up to 5 cm s^{-1} , from 5 to 10 cm s^{-1} , etc.) minus the average neuronal activity at rest, and then divided by their sum. This index weights how much the firing rate at a specific speed is stronger than the activity at rest. It is close to 1 when the firing rate at that given speed is markedly different to the baseline.

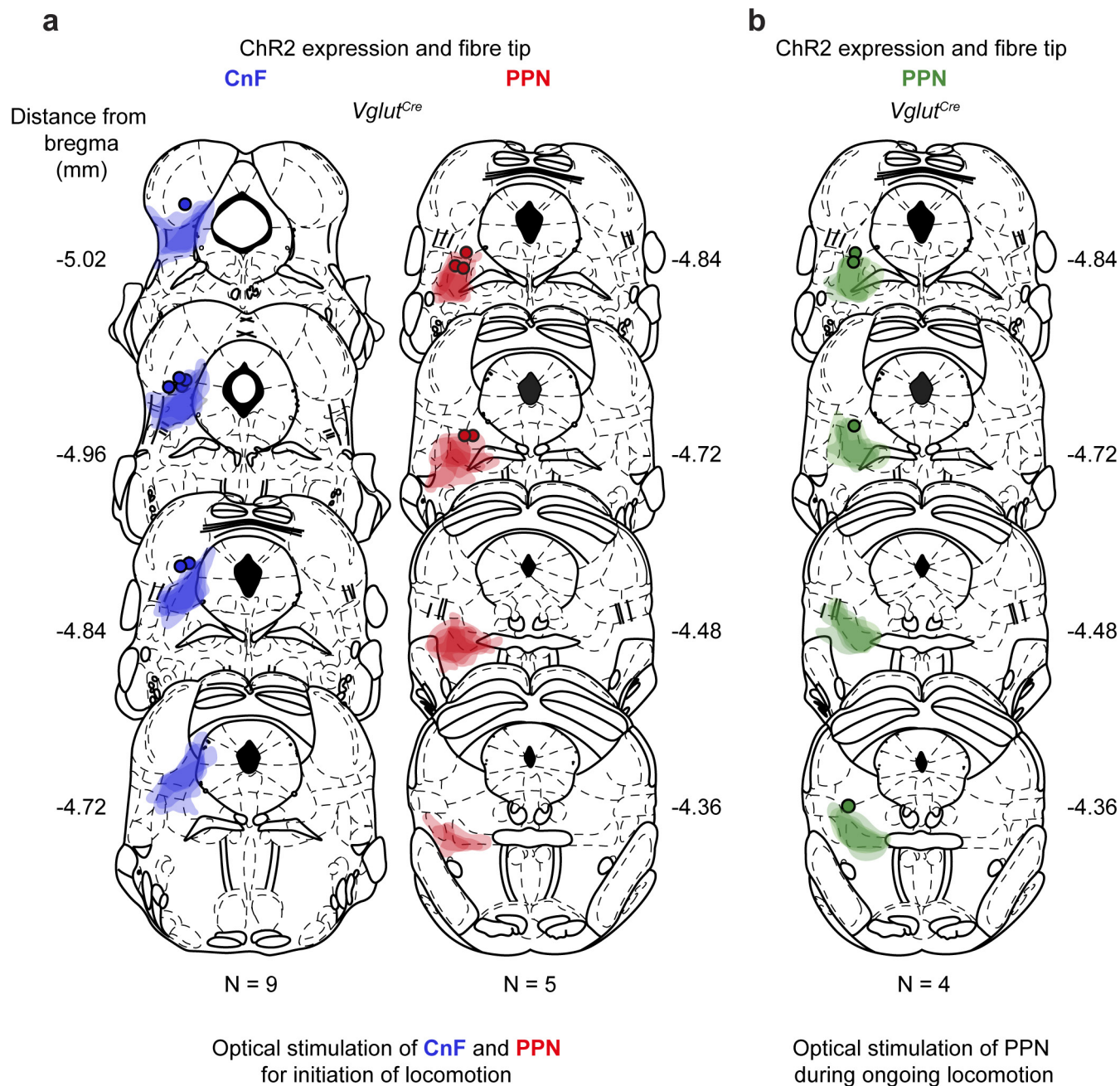
Tracking in hole-board. Head-dipping behaviour was recorded using a camera (30 frames per second) placed above the test box. Average speed, distance moved and duration of the head dips were measured using Ethovision software (Noldus Information Technology Inc.). The total number of head dips (hole visits) for each single hole was corrected by visual inspection of an experimenter blind to group and treatment. For optogenetically induced exploration, data were collected in 10 s stimulus periods. Only trials in which mice were exploring for less than 25% of the time before light stimulation were included in the analysis, to avoid behavioural adaptation.

Data availability. The datasets generated and/or analysed during the current study are available from the corresponding author upon reasonable request.

Code availability. Code used for analysis is available from the corresponding author upon reasonable request.

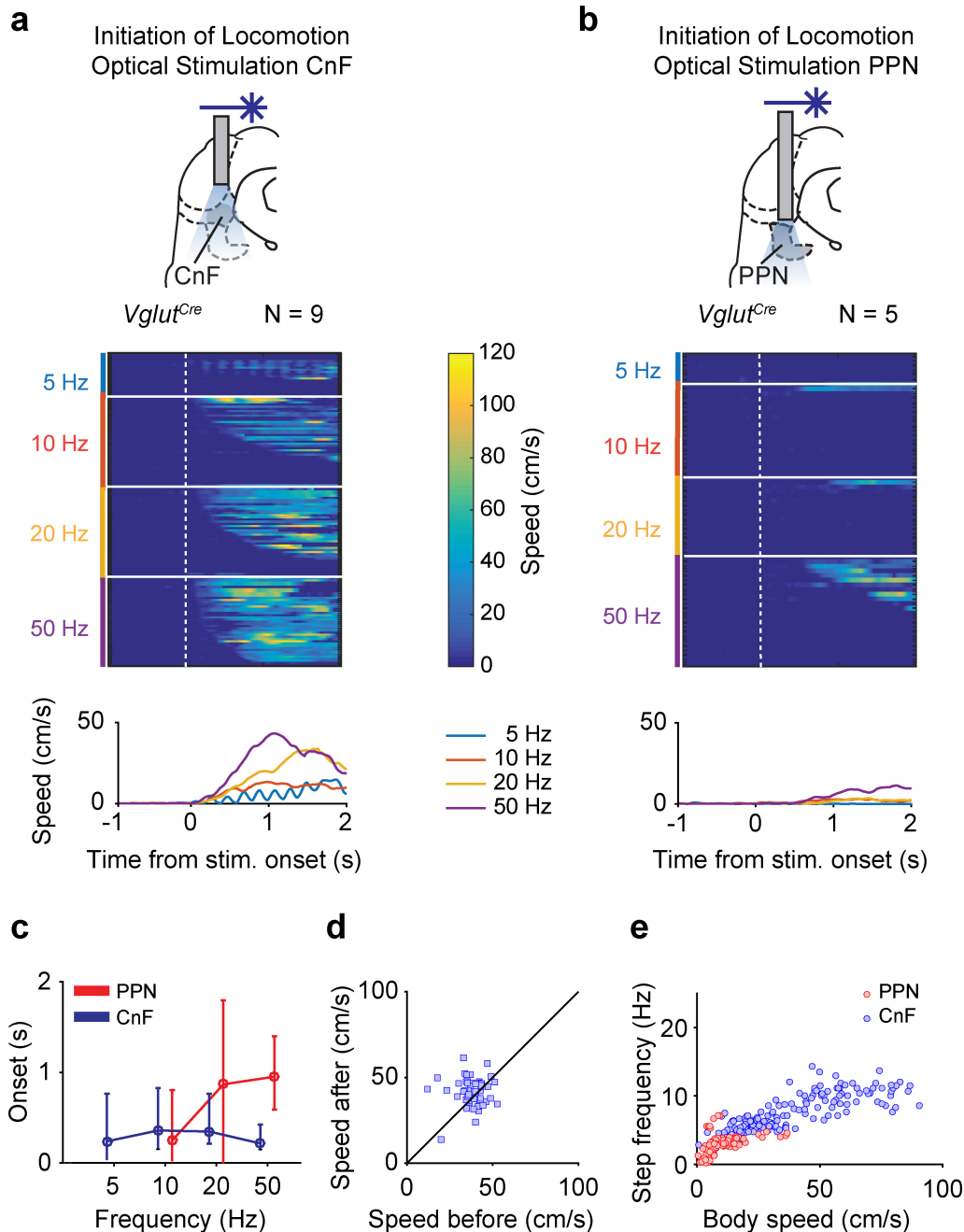
Statistics. Throughout the paper, the level of significance is indicated as * for $P < 0.05$, ** for $P < 0.01$ and *** for $P < 0.005$. All statistical tests used were two-tailed. Exact P values less than 0.001 were reported as $P < 0.001$. Non-parametric Kruskal–Wallis tests were used for non-matched data, and Friedman tests were used for repeated measurements. Correction for multiple comparisons was performed using the Bonferroni method. Custom scripts in MATLAB or R were used for the generation of graphs and statistical measurements. Wherever reported, data are medians and error bars indicate the 25th and 75th percentiles of the distribution, unless specified otherwise.

39. Bouvier, J. *et al.* Descending command neurons in the brainstem that halt locomotion. *Cell* **163**, 1191–1203 (2015).
40. Tovote, P. *et al.* Midbrain circuits for defensive behaviour. *Nature* **534**, 206–212 (2016).
41. Quiroga, R. Q., Nadasdy, Z. & Ben-Shaul, Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput.* **16**, 1661–1687 (2004).



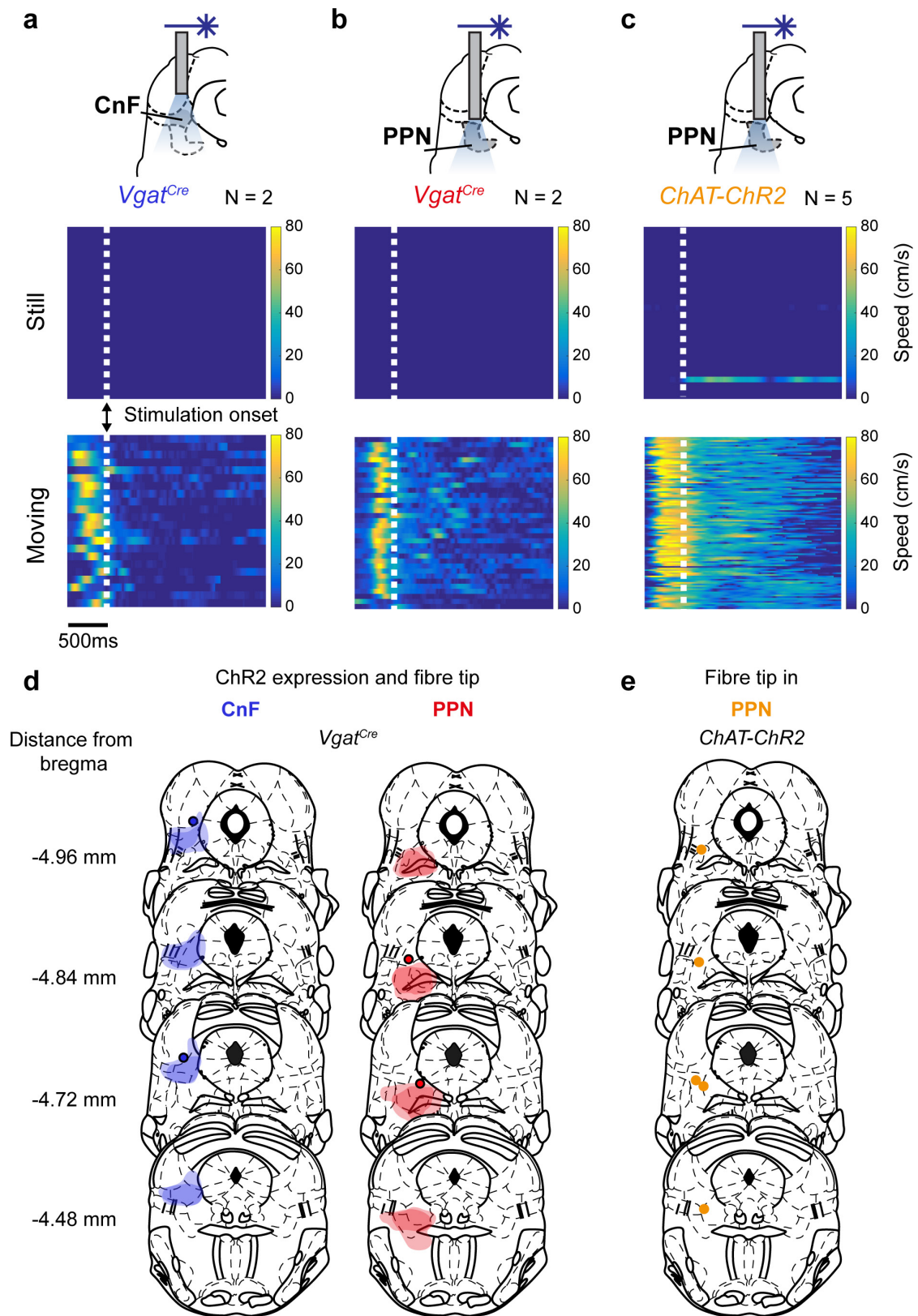
Extended Data Figure 1 | ChR2 expression in the CnF and the PPN. This figure summarizes the behavioural data in Fig. 1 and Extended Data Fig. 2a, b. **a**, Expression of ChR2 and fibre-tip positions in the CnF (left) and the PPN (right) for the data in Fig. 1 and Extended Data Fig. 2a–c, e. Coronal brain sections with viral expression from injected *Vglut2^{Cre}* mice were superimposed on sections redrawn from a mouse brain atlas³⁸.

The darker contour colours indicate the centre of expression, whereas the lighter colours indicate the border of the most extended expression. The round dots show the tip of the fibre. **b**, Expression of ChR2 and fibre-tip positions for the PPN data in Extended Data Fig. 2d. The mouse brain schematics in this figure have been reproduced with permission from Elsevier³⁸.



Extended Data Figure 2 | Control of locomotion speed from glutamatergic neurons in the CnF and the PPN. **a, b**, Speed profiles of mice after the stimulation of *Vglut2⁺ChR2* CnF (**a**) and *Vglut2⁺ChR2* PPN (**b**) neurons. Top panels show the location of optical stimulation in the CnF (**a**) and the PPN (**b**). Middle panels show colour plots of individual trials after the stimulation of *Vglut2⁺ChR2* CnF (**a**) and *Vglut2⁺ChR2* PPN (**b**) neurons (Fig. 1). The x axis represents time and the y axis represents trials at different stimulation frequencies. Data are aligned to the onset of stimulation (stim.). The colour gradient illustrates speed, with dark blue representing no movement and colours towards yellow representing the increase in speed (up to 120 cm s^{-1})

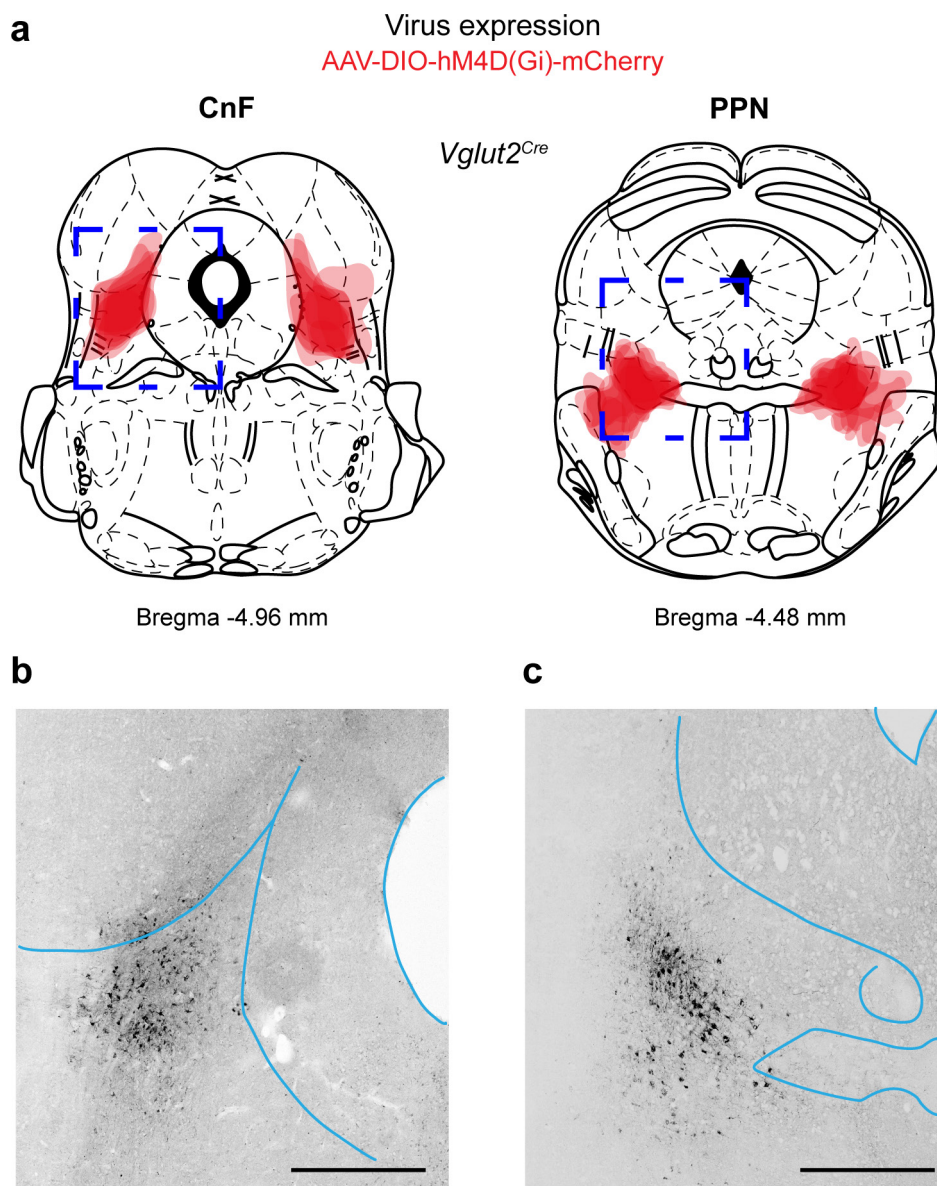
of the mouse in the linear corridor. Bottom panels show speed profiles obtained as an average of the movements at each stimulation frequency. **c**, Latencies to onset of locomotion from the stimulation of *Vglut2⁺ChR2* PPN (red) and *Vglut2⁺ChR2* CnF (blue) neurons as a function of the stimulation frequency. Error bars indicate the 25th and 75th percentiles of the distribution. **d**, Post-stimulus locomotor speed plotted against pre-stimulus locomotor speed in *Vglut2^{Cre}* mice that had been injected in the PPN with AAV-DIO-ChR2-mCherry ($n = 50$ trials from $N = 4$ mice). **e**, Step frequency plotted against speed of locomotion for the stimulation of *Vglut2⁺ChR2* PPN neurons (red, $n = 84$ trials from $N = 5$ mice) or *Vglut2⁺ChR2* CnF neurons (blue, $n = 173$ trials from $N = 9$ mice).



Extended Data Figure 3 | See next page for caption.

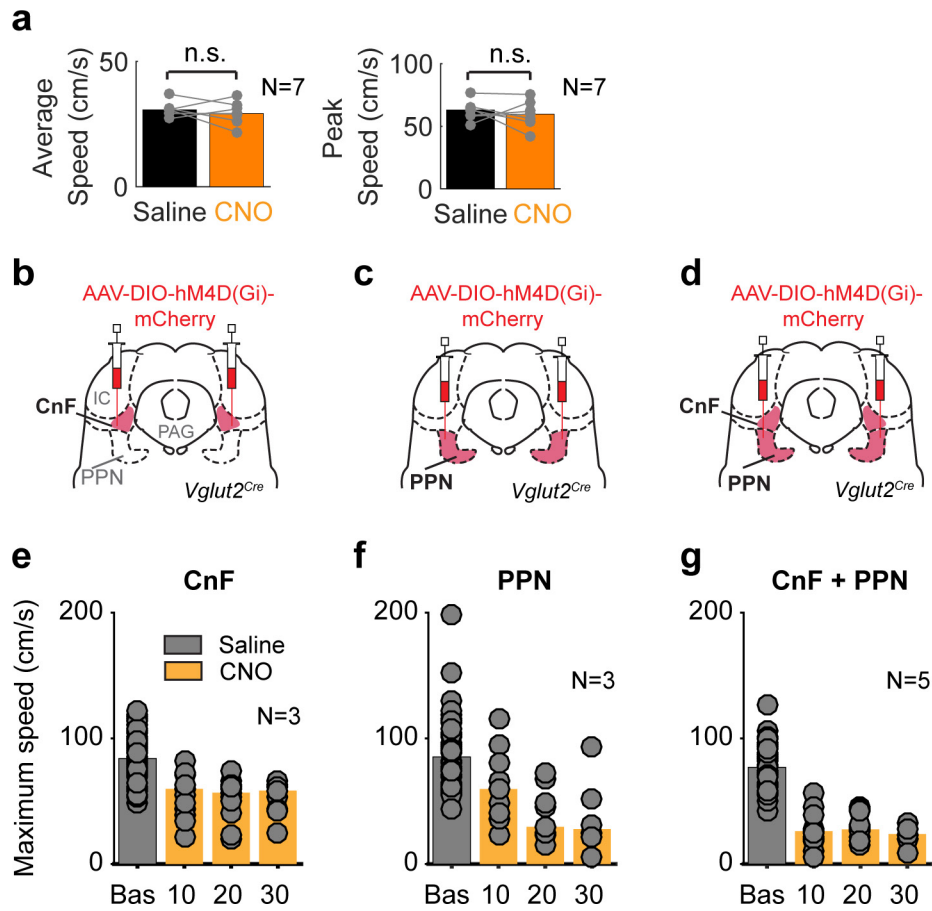
Extended Data Figure 3 | Activation of inhibitory neurons in the CnF or the PPN, and cholinergic neurons in the PPN, does not initiate locomotion but may modulate ongoing locomotion. **a–c,** Top panels show the implantation of the optical fibre to stimulate inhibitory cells in the CnF (**a**) and the PPN (**b**), and the cholinergic cells in the PPN (**c**). AAV-DIO-ChR2 virus was injected in *Vgat^{cre}* mice to target inhibitory cells, whereas cholinergic neurons expressed ChR2 transgenically by crossing *Chat^{cre}* with *RC26-ChR2^{flx/flx}* mice. Experiments were performed 3–4 weeks after injection of the virus, with mice locomoting spontaneously in a linear corridor. Middle and bottom panels show colour plots in which the *x* axis represents time and the *y* axis represents different trials, when the mice were not locomoting (middle panels, ‘still’) or when they were locomoting (bottom panels, ‘moving’) before the stimulation. Data are aligned to the onset of stimulation (dotted lines). The colour gradient illustrates speed, with dark blue representing no movement and colours

towards yellow representing an increase in speed (up to 60–80 cm s^{−1}) of the mouse in the linear corridor. Speed before versus after stimulation: CnF-*Vgat* inhibitory neurons: from still, $P > 0.05$, Wilcoxon signed-rank test (two sided) ($n = 18$, $N = 2$); when moving, from 27.9 cm s^{−1} to 4.2 cm s^{−1} $P < 0.05$, Wilcoxon signed-rank test ($n = 22$, $N = 2$). PPN-*Vgat* inhibitory neurons: from still, $P > 0.05$ ($n = 5$, $N = 2$); when moving from 27.6 cm s^{−1} to 8.6 cm s^{−1}, $P < 0.05$ Wilcoxon signed-rank test (two-sided) ($n = 34$, $N = 2$). Stimulation of long-projecting cholinergic cells in the PPN: from still, $P > 0.05$, Wilcoxon signed-rank test ($n = 102$, $N = 5$); when moving: before 47.3 cm s^{−1}, after 22.9 cm s^{−1}, $P < 0.05$, Wilcoxon signed-rank test (two-sided) ($n = 88$, $N = 5$). *n*, number of trials; *N*, number of mice. **d,** Diagram of viral expression and fibre-tip positions in *Vgat^{cre}* mice in the CnF (left) and the PPN (right). **e,** Diagram of fibre-tip positions in *Chat^{cre}* mice. The mouse brain schematics in this figure have been reproduced with permission from Elsevier³⁸.



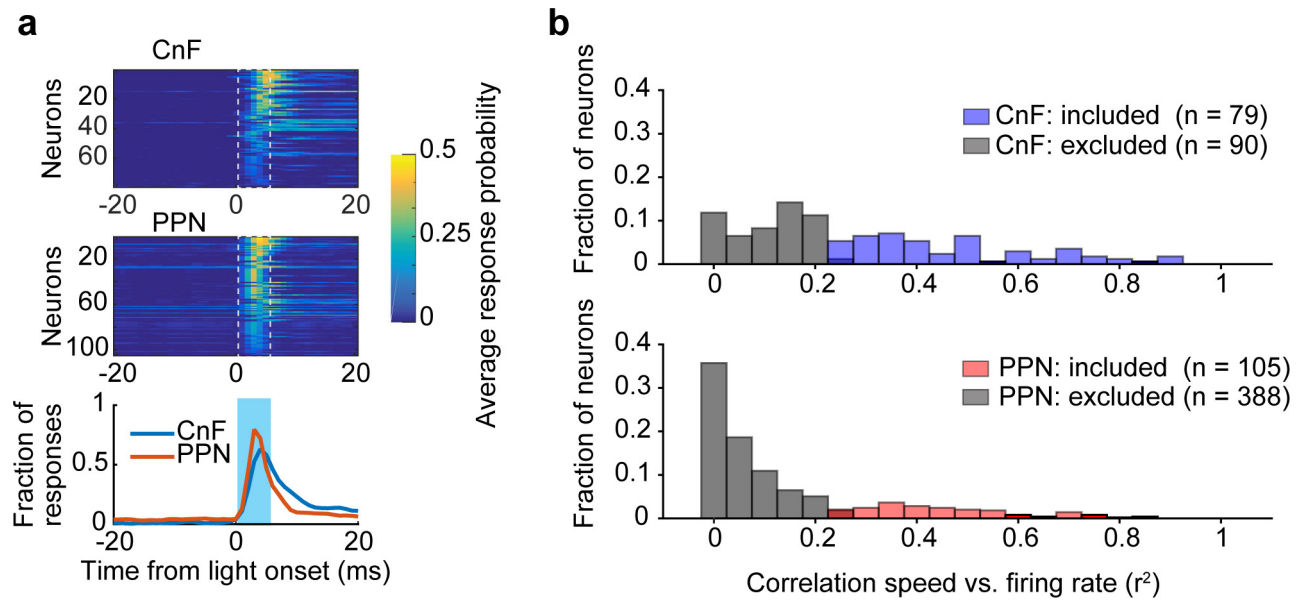
Extended Data Figure 4 | Summary diagram of iDREADD injection sites in the CnF and the PPN. **a**, Expression of iDREADD in *Vglut2*⁺ neurons of the CnF (left, *N* = 8) or the PPN (right, *N* = 9) in mice used in Fig. 2. **b**, **c**, Coronal sections showing the expression pattern of iDREADD

in *Vglut2*⁺CnF (**b**) and *Vglut2*⁺PPN (**c**) neurons. Scale bars, 500 μ m. The mouse brain schematics in this figure have been reproduced with permission from Elsevier³⁸.



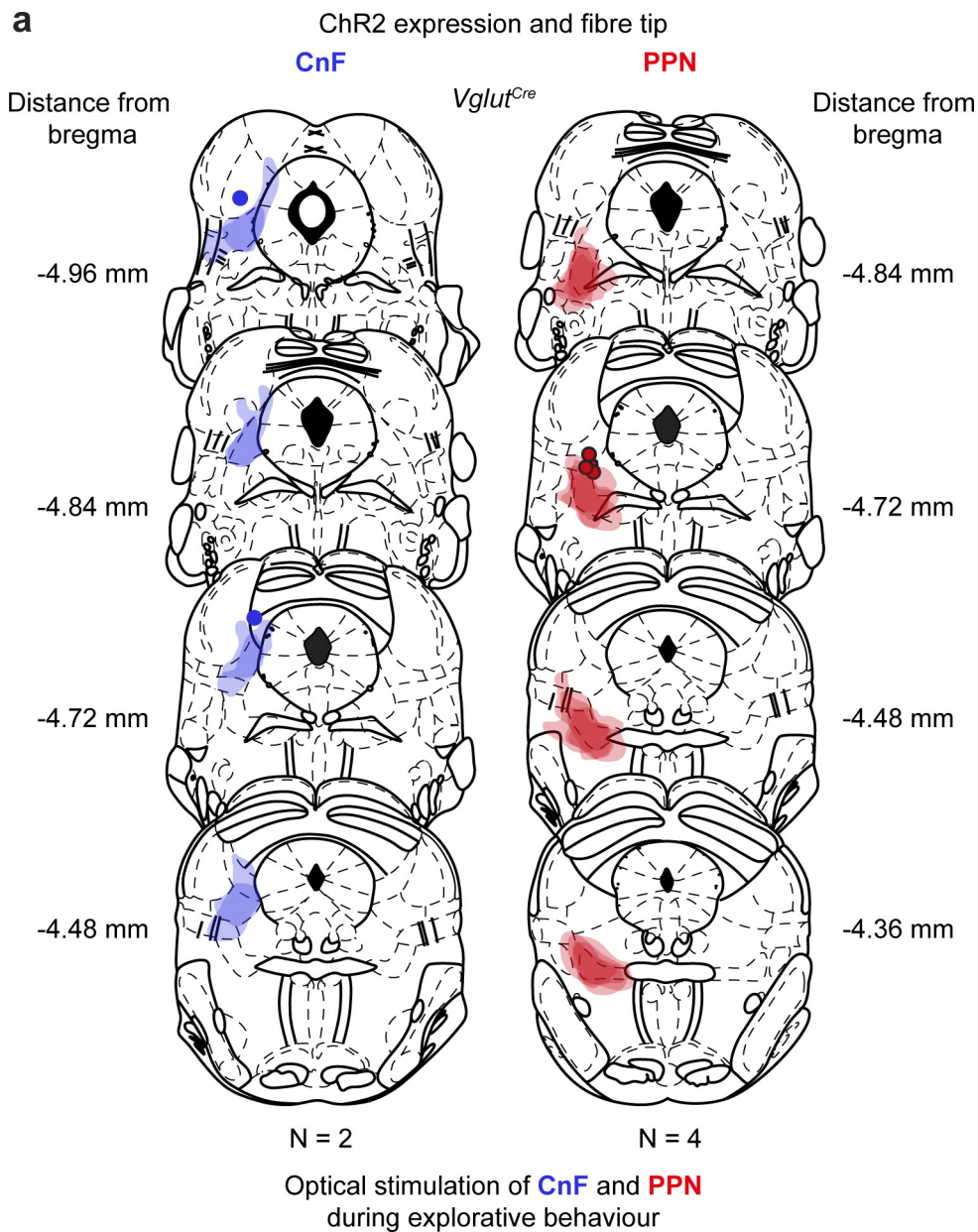
Extended Data Figure 5 | Control for CNO injection and time course of the silencing effect of glutamatergic neurons in the CnF and the PPN. **a**, Average (left) and maximum (right) speeds attained by wild-type mice during treadmill experiments after the intraperitoneal injection of saline (black) and CNO (orange, 1 mg kg⁻¹) (N=7). There was no significant difference in these speed parameters between the saline and CNO experiments (Wilcoxon signed-rank, two-sided, $P > 0.45$). **b–d**, Sites of

AAV-DIO-hM4D(Gi)-mCherry injection in *Vglut2^{Cre}* mice in the CnF (**b**), the PPN (**c**) or the CnF and PPN (**d**). CNO was injected intraperitoneally and locomotor performance was tested on a treadmill. **e–g**, Graphs show the development of the inhibition of glutamatergic cells in the CnF (**e**, N=3), the PPN (**f**, N=3) or the CnF and PPN (**g**, N=5) on maximal locomotor speed over time. Grey bars, baseline. Orange bars, time (in min) after CNO administration. Points show individual trials.



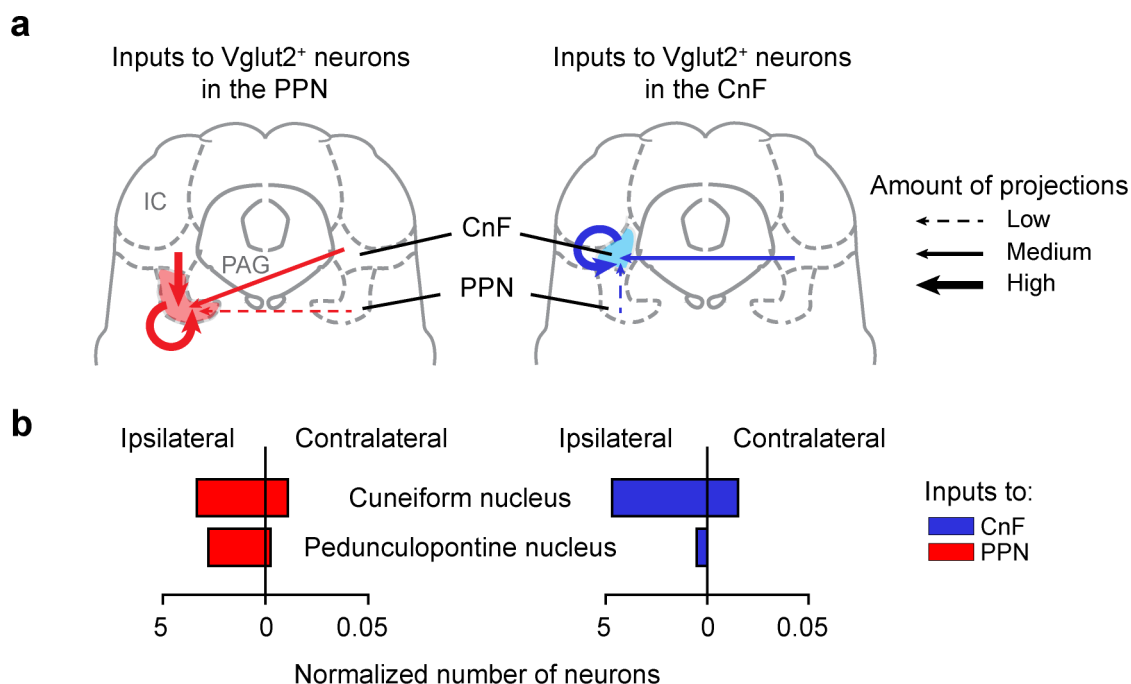
Extended Data Figure 6 | Latencies of the light activation of PPN and CnF neurons, and fractions of Vglut2⁺ChR2 neurons of the CnF and PPN with speed-related activity. **a**, Latencies of light activation of all cells included in the analysis. **b**, Distribution of Vglut2⁺ChR2 CnF neurons (blue bars, $n = 79$ out of 169) and Vglut2⁺ChR2 PPN neurons (red bars,

$n = 105$ out of 493) showing the correlation of firing activity with the locomotor speed of the mouse. In both panels, grey bars indicate neurons that show no significant correlation with the speed (Spearman correlation test, $P > 0.05$).



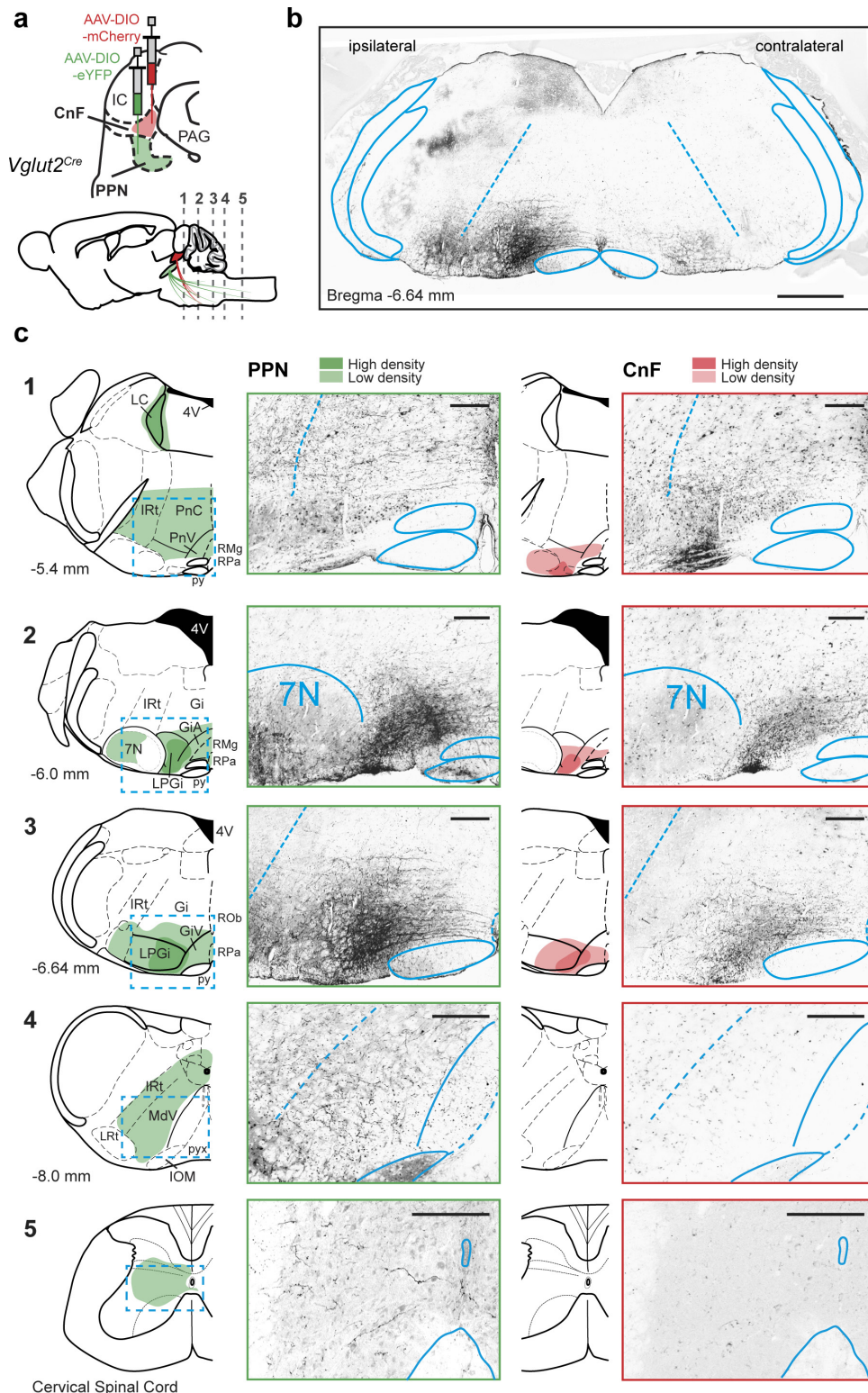
Extended Data Figure 7 | Summary of injection sites in the PPN and the CnF for hole-board stimulation experiments. a, Expression of ChR2 and fibre-tip positions in the CnF (left) or the PPN (right) for mice used in the

experiments shown in Fig. 5d, e. The mouse brain schematics in this figure have been reproduced with permission from Elsevier³⁸.



Extended Data Figure 8 | Connectivity between the PPN and the CnF.
a, b, AAV-EF1a-FLEX-GTB helper virus followed by EnvA G-deleted-rabies-mCherry virus were unilaterally injected in the PPN (left, red) or the CnF (right, blue) in *Vglut2^{cre}* mice to trace inputs to glutamatergic neurons. Schematics summarizing the inputs to Vglut2⁺ PPN neurons

(red) and Vglut2⁺CnF neurons (blue) are shown in **a**; the thickness of the arrows indicates the amount of connectivity based on the counts of the normalized number of neurons as shown in **b**. Dashed arrows indicate sparse connectivity.



Extended Data Figure 9 | The CnF and PPN have different descending output matrices. **a**, Simultaneous unilateral injection (top) of AAV-DIO-ChR2 virus in the CnF (mCherry, red) and the PPN (eYFP, green) in *Vglut2^{Cre}* mice ($N = 3$). Sagittal view of the brain (bottom) displaying the location in the brainstem (1–4) and the spinal cord (5) of the coronal sections shown in **c**. **b**, Coronal section showing ipsilateral and contralateral projection areas from glutamatergic PPN neurons. **c1–5**, Schematics and coronal sections showing projection areas from glutamatergic PPN (left, green) and CnF (right, red) neurons onto nuclei in the pons, medulla and spinal cord. In the schematics, the darker shades delineate the areas with the highest density of projections. In coronal sections, labelled processes are seen in black. Anatomical landmarks are

indicated in the schematics. Scale bars, 200 μm . 4V, fourth ventricle; 7N, facial motor nucleus; Gi, gigantocellular nucleus; GiA, gigantocellular reticular nucleus, alpha part; GiV, gigantocellular reticular nucleus, ventral part; IOM, inferior olive, medial nucleus; IRT, intermediate reticular nucleus; LC, locus coeruleus; LPGi, lateral paragigantocellular nucleus; LRT, lateral reticular nucleus; MdV, medullary reticular nucleus, ventral part; PnC, pontine reticular nucleus, caudal part; PnV, pontine reticular nucleus, ventral part; py, pyramidal tract; pyx, pyramidal decussation; RMg, raphe magnus; ROb, raphe obscurus; RPa, raphe pallidus. The mouse brain schematics in this figure have been reproduced with permission from Elsevier³⁸.

α -Klotho is a non-enzymatic molecular scaffold for FGF23 hormone signalling

Gaozhi Chen^{1,2*}, Yang Liu^{2*}, Regina Goetz², Lili Fu^{1,2}, Seetharaman Jayaraman³, Ming-Chang Hu⁴, Orson W. Moe⁴, Guang Liang¹, Xiaokun Li¹ & Moosa Mohammadi²

The ageing suppressor α -klotho binds to the fibroblast growth factor receptor (FGFR). This commits FGFR to respond to FGF23, a key hormone in the regulation of mineral ion and vitamin D homeostasis. The role and mechanism of this co-receptor are unknown. Here we present the atomic structure of a 1:1:1 ternary complex that consists of the shed extracellular domain of α -klotho, the FGFR1c ligand-binding domain, and FGF23. In this complex, α -klotho simultaneously tethers FGFR1c by its D3 domain and FGF23 by its C-terminal tail, thus implementing FGF23–FGFR1c proximity and conferring stability. Dimerization of the stabilized ternary complexes and receptor activation remain dependent on the binding of heparan sulfate, a mandatory cofactor of paracrine FGF signalling. The structure of α -klotho is incompatible with its purported glycosidase activity. Thus, shed α -klotho functions as an on-demand non-enzymatic scaffold protein that promotes FGF23 signalling.

Endocrine FGF23 regulates phosphate and vitamin D homeostasis by reducing the cell surface expression of sodium phosphate co-transporters and by repressing transcription of rate-limiting enzymes for vitamin D biosynthesis^{1,2} in the kidney. FGF23 exerts its metabolic functions by binding and activating FGFR tyrosine kinases³ in an α -klotho co-receptor dependent fashion. The extracellular domain of a prototypical FGFR consists of three immunoglobulin-like domains: D1, D2 and D3. The membrane proximal portion comprising D2, D3 and the D2–D3 linker (FGFR^{ecto}) is both necessary and sufficient for FGF ligand binding^{4,5}. Tissue-specific alternative splicing in the D3 domain of FGFR1–FGFR3 generates ‘b’ and ‘c’ isoforms, each with distinct ligand-binding specificity^{5,6}. α -klotho, fortuitously discovered as an ageing-suppressor gene⁷, is a single-pass transmembrane protein with an extracellular domain composed of two tandem domains (KL1 and KL2), each with notable homology to family 1 glycosidases⁸ (Extended Data Fig. 1a). Membrane-bound α -klotho (α -klothoTM) associates with cognate FGFRs of FGF23, namely the ‘c’ splice isoforms of FGFR1 and FGFR3 (FGFR1c and FGFR3c) and FGFR4^{9–12}. This enables them to bind and respond to FGF23^{9,11,12}. α -klothoTM is predominantly expressed in the kidney distal tubules, the parathyroid gland, and the brain choroid plexus^{7,13}, and this is considered to determine the target tissue specificity of FGF23^{11,12}. Cleavage of α -klothoTM by ADAM proteases^{14,15} in kidney distal tubules sheds the α -klotho ectodomain (α -klotho^{ecto}; Extended Data Fig. 1a) into body fluids, for example, serum, urine and cerebrospinal fluid^{16–19}. α -Klotho^{ecto} is thought to lack co-receptor activity and act as a circulating anti-ageing hormone independently of FGF23^{20,21}. A plethora of activities has been attributed to shed α -klotho^{ecto}, the bulk of which require a purported intrinsic glycosidase activity^{22–25}.

Here we show that circulating α -klotho^{ecto} is an on-demand bona fide co-receptor for FGF23, and determine its crystal structure in complex with FGFR1c^{ecto} and FGF23. The structure reveals that

α -klotho serves as a non-enzymatic scaffold that simultaneously tethers FGFR1c and FGF23 to implement FGF23–FGFR1c proximity and hence stability. Surprisingly, heparan sulfate (HS), a mandatory cofactor for paracrine FGFs, is still required as an ancillary cofactor to promote the formation of a symmetric 2:2:2:2 FGF23–FGFR1c–Klotho–HS quaternary signalling complex.

Soluble α -klotho^{ecto} acts as a co-receptor for FGF23

To determine whether soluble α -klotho^{ecto} can support FGF23 signalling, α -klotho-deficient HEK293 cells, which naturally express FGFRs, were incubated with a concentration of α -klotho^{ecto} sufficient to drive all available cell-surface cognate FGFRs into binary complexed form. After brief rinses with PBS, the cells were stimulated with increasing concentrations of FGF23. In parallel, a HEK293 cell line that overexpresses membrane-bound α -klotho (HEK293- α -klothoTM) was treated with increasing concentrations of FGF23. The dose-response for FGF23-induced ERK phosphorylation in α -klotho^{ecto}-pretreated untransfected HEK293 cells was similar to that observed in HEK293- α -klothoTM cells (Extended Data Fig. 1b, top), suggesting that α -klotho^{ecto} can serve as a co-receptor for FGF23. Pre-treatment of HEK293- α -klothoTM cells with α -klotho^{ecto} did not result in any further increase in FGF23 signalling, indicating that all cell-surface FGFRs in this cell line were in binary FGFR- α -klothoTM form (Extended Data Fig. 1b, bottom). We conclude that soluble and transmembrane forms of α -klotho possess a similar capacity to support FGF23 signalling. Consistent with these results, injection of wild-type mice with α -klotho^{ecto} protein led to an increase in renal phosphate excretion and a decrease in serum phosphate (Extended Data Fig. 1c). Notably, it also led to a 1.5-fold increase in *Egr1* transcripts in the kidney (Extended Data Fig. 1d), demonstrating that α -klotho^{ecto} can serve as a bona fide co-receptor to support FGF23 signalling in renal proximal tubules. In light of these data, we propose that the pleiotropic anti-ageing effects of α -klotho are all dependent on FGF23.

¹Chemical Biology Research Center, School of Pharmaceutical Sciences, Wenzhou Medical University, Wenzhou, Zhejiang 325035, China. ²Department of Biochemistry & Molecular Pharmacology, New York University School of Medicine, New York, New York 10016, USA. ³New York Structural Biology Center, New York, New York 10027, USA. ⁴Departments of Internal Medicine and Physiology, and Charles and Jane Pak Center of Mineral Metabolism and Clinical Research, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.

*These authors contributed equally to this work.

Structural basis of α -klotho co-receptor function

We solved the crystal structure of a human 1:1:1 FGF23–FGFR1c^{ecto}– α -klotho^{ecto} ternary complex at 3.0 Å resolution (Extended Data Table 1). In this complex, α -klotho^{ecto} serves as a massive scaffold, tethering both FGFR1c and FGF23 to itself. In doing so, α -klotho^{ecto} enforces FGF23–FGFR1c proximity and thus augments FGF23–FGFR1c binding affinity (Fig. 1). The overall geometry of the ternary complex is compatible with its formation on the cell surface (Extended Data Fig. 2a).

The binary FGF23–FGFR1c^{ecto} complex adopts a canonical FGF–FGFR complex topology, in which FGF23 is bound between the D2 and D3 domains of the receptor, engaging both these domains and a short interdomain linker (Extended Data Fig. 3a). However, compared to paracrine FGFs, FGF23 makes fewer or weaker contacts with the D3 domain and D2–D3 linker, explaining the inherently low affinity of FGF23 for FGFR1c (Extended Data Fig. 3b, c). Notably, analysis of the binding interface between FGF23 and FGFR1c D3 in the crystal structure reveals specific contacts between FGF23 and a serine residue uniquely present in the ‘c’ splice isoforms of FGFR1–FGFR3 and in FGFR4 (Extended Data Fig. 4a). Indeed, replacing this ‘c’-isoform-specific serine residue with a ‘b’-isoform-specific tyrosine impaired FGF23 signalling (Extended Data Fig. 4b, c). We conclude that the FGF binding specificity inherent to FGF23 operates alongside that of α -klotho (Extended Data Fig. 4d, e) to restrict FGF23 signalling to the ‘c’ splice isoforms and FGFR4^{11,12}.

In the ternary complex, α -klotho^{ecto} exists in an extended conformation. Consistent with their sequence homology to the glycoside hydrolase A clan⁸, the α -klotho KL1 (Glu34 to Phe506) and KL2 (Leu515 to Ser950) domains each assume a ($\beta\alpha$)₈ triosephosphate isomerase (TIM) barrel fold consisting of an inner eight-stranded parallel β -barrel and eight surrounding α -helices (Fig. 2a and Extended Data Fig. 5a). The two KL domains are connected by a short, proline-rich and hence stiff linker (Pro507 to Pro514) (Fig. 1a, b). KL1 sits atop KL2, engaging it via a few interdomain contacts involving the N terminus preceding the β 1 strand and the α 7 helix of KL1, and the β 5 α 5 and β 6 α 6 loops and α 7 helix of KL2 (Extended Data Fig. 2b). Notably, one of the interdomain contacts is mediated by a Zn²⁺ ion (Fig. 3c and Extended Data Fig. 2b, c). These contacts stabilize the observed elongated conformation of α -klotho^{ecto}, creating a deep cleft between the two KL domains. This merges with a wide-open central β -barrel cavity in KL2, and forms a large binding pocket that tethers the distal C-terminal tail of FGF23 past the 176-Arg-His-Thr-Arg-179 proteolytic cleavage site (Fig. 1b). Meanwhile, the long β 1 α 1 loop of KL2 (Fig. 2a) protrudes as much as 35 Å away from the KL2 core to latch onto the FGFR1c D3 domain, thus anchoring the receptor to α -klotho (Fig. 1b). Accordingly, we have named this KL2 loop the receptor binding arm (RBA; residues 530–578; Extended Data Fig. 5a).

We superimposed the TIM barrels of KL1 and KL2 onto that of klotho-related protein (KLrP, also known as GBA3), the cytosolic member of the klotho family with proven glycosylceramidase activity²⁶. This comparison revealed major conformational differences in the loops surrounding the entrance to the catalytic pocket in KL1 and KL2 (Fig. 2b and Extended Data Fig. 5b–d). Moreover, both KL domains lack one of the key catalytic glutamates deep within the putative catalytic pocket. These substantial differences are incompatible with an intrinsic glycosidase activity for α -klotho^{ecto}. Indeed, α -klotho^{ecto} failed to hydrolyse substrates for both sialidase and β -glucuronidase *in vitro* (Fig. 2c). Together, our data define α -klotho as the only known example of a TIM barrel protein that serves purely as a non-enzymatic molecular scaffold.

Binding interface between α -klotho and FGFR1c

The interface between α -klotho RBA and FGFR1c D3 (Fig. 3a) buries over 2,200 Å² of solvent-exposed surface area, which is consistent with the high affinity of α -klotho binding to FGFR1c (dissociation constant (K_d) = 72 nM)¹⁰. At the distal tip of the RBA, residues 547–Tyr-Leu-Trp-549 and 556–Ile-Leu-Arg-558 form a short β -strand pair

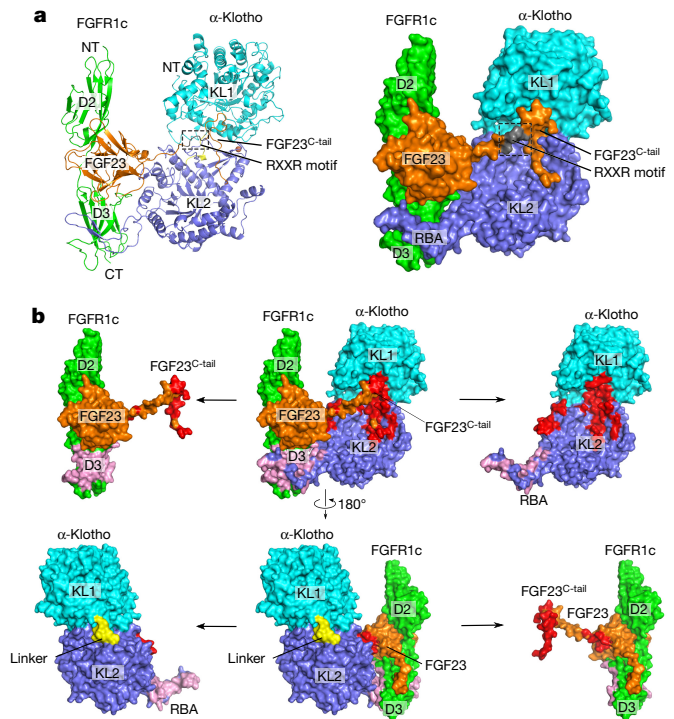


Figure 1 | Overall topology of the FGF23–FGFR1c^{ecto}– α -klotho^{ecto} complex. **a**, Cartoon (left) and surface representation (right) of the ternary complex structure. The α -klotho KL1 (cyan) and KL2 (blue) domains are joined by a short proline-rich linker (yellow; not visible in the surface presentation). FGF23 is in orange with its proteolytic cleavage motif in grey. FGFR1c is in green. CT, C terminus; NT, N terminus. **b**, Binding interfaces between α -klotho^{ecto} and the FGF23–FGFR1c^{ecto} complex. The ternary complex (centre) is shown in two different orientations related by a 180° rotation along the vertical axis. FGF23– α -klotho^{ecto} (red) and FGFR1c^{ecto}– α -klotho^{ecto} (pink) interfaces are visualized by pulling α -klotho^{ecto} and the FGF23–FGFR1c^{ecto} complex away from each other. The separated components are shown to the left and right of the ternary complex.

(RBA- β 1:RBA- β 2) as their hydrophobic side chains are immersed in a wide hydrophobic groove between the four-stranded β C′– β C– β F– β G sheet and the β C– β C′ loop of FGFR1c D3 (Fig. 3b, top). The RBA- β 1:RBA- β 2 strand pair forms an extended β -sheet with the β C′– β C– β F– β G sheet of D3 as the backbone atoms of RBA- β 1 and D3 β C′ make three hydrogen bonds that further augment the interface (Fig. 3b, bottom). Residues at the proximal end of the RBA engage a second smaller binding pocket at the bottom edge of D3 next to the hydrophobic groove (Extended Data Fig. 6a, b). Both α -klotho binding pockets in the receptor D3 domain differ between ‘b’ and ‘c’ splice isoforms. Leu342, for example, is strictly conserved in the ‘c’ splice isoforms of FGFR1–FGFR3 and FGFR4. This explains the previously described binding selectivity of α -klotho for this subset of FGFRs^{9,11,12} (Extended Data Fig. 4a).

Consistent with the crystal structure, soluble α -klotho lacking the RBA (α -klotho^{ecto/ Δ RBA}) failed to form a binary complex with FGFR1c^{ecto} in solution (Fig. 4a) and hence could not support FGF23 signalling (Fig. 4b). Likewise, membrane-bound α -klotho lacking the RBA (α -klotho^{TM/ Δ RBA}) was also disabled in acting as a FGF23 co-receptor (Fig. 4b). Importantly, α -klotho^{ecto/ Δ RBA} did not exhibit any phosphaturic activity *in vivo* (Extended Data Fig. 7a). On the contrary, the α -klotho^{ecto/ Δ RBA} mutant antagonized the activity of native α -klotho by sequestering FGF23 into functionally inactive binary complexes, that is, by acting as an FGF23 ligand trap (Extended Data Fig. 7). These data refute the concept that α -klotho^{ecto} functions as an FGF23-independent phosphaturic enzyme²⁴. Our conclusion is supported by a gene knockout study that compared the phenotypes of mice with knockout of FGF23 (*Fgf23*^{−/−}), mice with knockout of α -klotho (*Kl*^{−/−}) and double-knockout mice (*Fgf23*^{−/−}*Kl*^{−/−})²⁷.

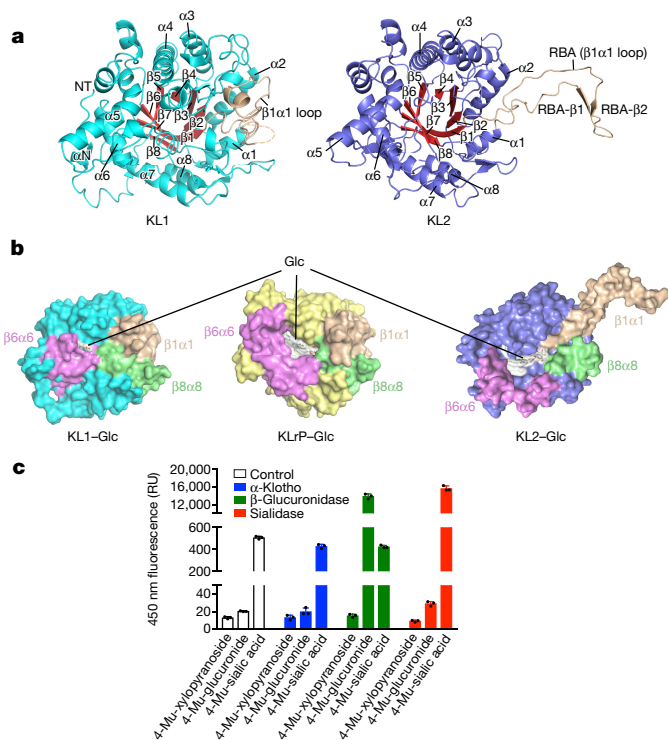


Figure 2 | α -Klotho is a non-enzymatic molecular scaffold.

a, Triosephosphate isomerase (TIM) barrel topology of the α -klotho KL1 and KL2 domains. KL1 is in the same orientation as in Fig. 1a, whereas KL2 has been superimposed onto KL1 and has thus been reoriented. The eight alternating β -strands (red) and α -helices (cyan/blue) that define the TIM barrel are labelled according to the standard nomenclature for the TIM fold⁸. KL1 and KL2 differ markedly in the conformation of the β 1 α 1 loop (wheat). In KL2, this loop protrudes away from the TIM barrel and serves as a receptor binding arm (RBA; Fig. 1). **b**, Molecular surfaces of KL1P-glucosylceramide (Glc) (centre; KL1P in yellow), KL1-Glc (left; KL1 in cyan) and KL2-Glc (right; KL2 in blue). Binding of Glc to KL1 and KL2 was simulated by superimposing KL1 and KL2 onto KL1P-Glc. In all cases, Glc is shown as pale grey sticks or surface. The divergent conformation of the β 6 α 6 loop (pink) in KL1 almost seals off the entrance to the catalytic pocket, while the divergent conformations of the β 1 α 1 (RBA; wheat), β 6 α 6 (pink) and β 8 α 8 (green) loops in KL2 leave the central barrel cavity in KL2 in a more solvent-exposed state that is less capable of ligating substrate (see also Extended Data Fig. 5). **c**, Glycosidase activity of α -klotho^{ecto}, sialidase and β -glucuronidase. Data are mean and s.d. Dots denote individual data points; $n = 3$ independent experiments. RU, relative units.

Binding interface between α -klotho and FGF23

Regions from both KL domains act together to recruit FGF23 (Fig. 1b), thus explaining why only an intact α -klotho ectodomain is capable of supporting FGF23 signalling^{12,28}. The interactions between FGF23 and α -klotho result in the burial of a large amount of solvent-exposed surface area (2,732 Å²), of which nearly two-thirds (1,961 Å²) are buried between the FGF23 C-terminal tail and α -klotho, and the remaining one-third is buried between the FGF23 core and α -klotho (Fig. 3a). At the interface between α -klotho and the FGF23 C-terminal tail, FGF23 residues 188–Asp–Pro–Leu–Asn–Val–Leu–193 adopt an unusual cage-like conformation (Fig. 3a, c), which is tethered by residues from both KL domains via hydrogen bonds and hydrophobic contacts deep inside the KL1–KL2 cleft (Fig. 3c). Further downstream, the side chains of Lys194, Arg196 and Arg198 of the FGF23 C-terminal tail dip into the central barrel cavity of KL2, making hydrogen bonds with several α -klotho residues (Fig. 3c). At the interface between the FGF23 β -trefoil core and α -klotho, residues from the β 5 β 6 turn and the α C helix of FGF23 make hydrogen bonds and hydrophobic contacts with residues in the short β 7 α 7 and β 8 α 8 loops at the upper rim of the KL2 cavity (Extended Data Fig. 6a, c).

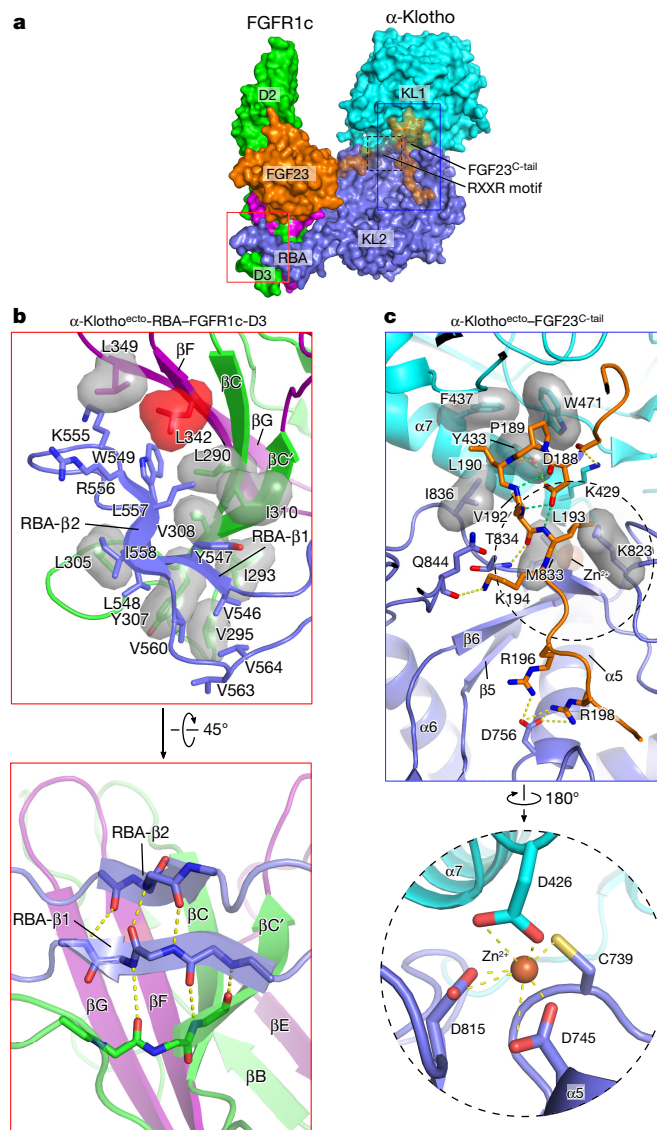


Figure 3 | α -Klotho simultaneously tethers FGFR1c by its D3 domain and FGF23 by its C-terminal tail. **a**, Ternary complex structure in surface representation. Colouring is as in Fig. 1a, except that the alternatively spliced region of FGFR1c is highlighted in purple. Red box denotes perimeter of interface between distal tip of α -klotho RBA and the hydrophobic FGFR1c D3 groove. Blue box denotes the perimeter of α -klotho–FGF23^{C-tail} interface. **b**, RBA stretches out of the KL2 domain of α -klotho^{ecto} and latches onto the FGFR1c D3 domain. Top, interface between the distal tip of RBA and the D3 groove detailing hydrophobic interactions (grey transparent surfaces). Note that Leu342 (red) from the spliced region of the D3 groove is strictly conserved in 'c' splice isoforms of FGFR1–FGFR3 and FGFR4 and is mutated in Kallmann syndrome³⁶. Bottom, close-up view of the extended β -sheet between the RBA- β 1:RBA- β 2 strand pair and the four-stranded β -sheet in D3 (β C'– β C– β F– β G). This structure forms via hydrogen bonding (dashed yellow lines) between backbone atoms of RBA- β 1 and D3- β C'. **c**, Both KL domains of α -klotho^{ecto} participate in tethering of the flexible C-terminal tail of FGF23 (FGF23^{C-tail}). FGF23^{C-tail} residues Asp188–Thr200 thread through the KL1–KL2 cleft and the β -barrel cavity of KL2. Of these residues, Asp188–Leu193 adopt a cage-like conformation that is partially stabilized by intramolecular hydrogen bonds (dashed green lines). Dashed yellow lines denote intermolecular hydrogen bonds; grey transparent surfaces denote hydrophobic interactions. Note that Tyr433 from the KL1 α 7 helix deep inside the KL1–KL2 cleft has a prominent role in tethering the cage-like structure in the FGF23^{C-tail} formed by Asp188–Leu193. Dashed circle (shown at greater magnification below) denotes the KL1–KL2 interface where residues from both α -klotho domains jointly coordinate a Zn²⁺ ion (orange sphere).

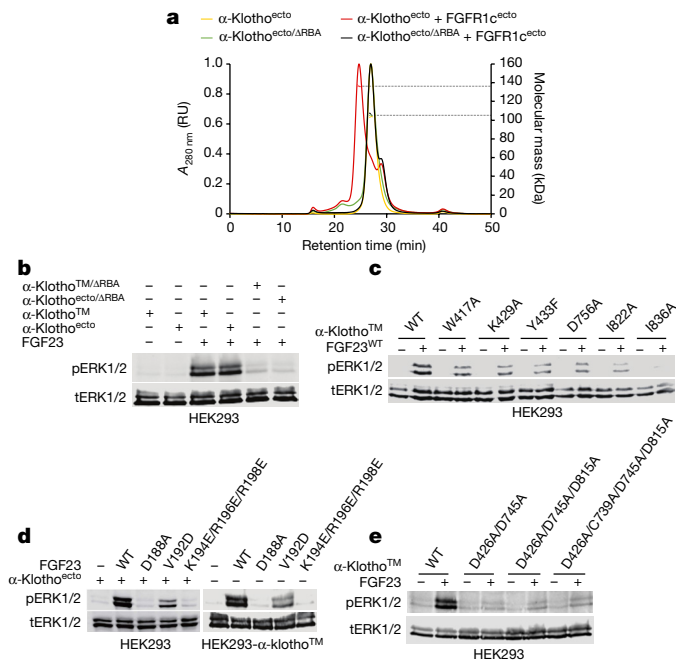


Figure 4 | Mutagenesis experiments validate the crystallographically deduced mode of ternary complex formation. **a**, Size exclusion chromatography–multi-angle light scattering (SEC–MALS) analysis of FGF23^{ecto} interaction with wild-type α-klotho^{ecto} or its RBA deletion mutant. **b–e**, Representative immunoblots of phosphorylated ERK (pERK1/2; top) and total ERK (tERK1/2; bottom; sample loading controls) in total HEK293 cell lysates ($n = 3$ independent experiments for each panel). **b**, Analysis of the effects of RBA deletion on the co-receptor activity of α-klotho^{ecto} and α-klothoTM isoforms. **c**, Analysis of mutations in the α-klotho binding pocket that engages the FGF23^{C-tail}. WT, wild type. **d**, Analysis of mutations in the FGF23^{C-tail} that disrupt α-klotho–FGF23^{C-tail} interaction. **e**, Analysis of mutations of the four Zn²⁺-coordinating amino acids in α-klotho.

To test the biological relevance of the observed contacts between α-klotho and the FGF23 C-terminal tail, we introduced several mutations into α-klothoTM and FGF23 to disrupt α-klotho–FGF23 binding (Fig. 4c). Consistent with our structure-based predictions, all α-klothoTM mutants showed an impaired ability to support FGF23 signalling (Fig. 4c). The FGF23 mutants also exhibited a reduced ability to signal, regardless of whether soluble or membrane-bound α-klotho served as co-receptor (Fig. 4d). Remarkably, the FGF23(D188A) mutant (which eliminates the intramolecular hydrogen bonds that support cage conformation) was totally inactive, underscoring the importance of the cage-like conformation in the tethering of FGF23 to α-klotho. Notably, tethering of this cage-like structure requires precise alignment of residues from both KL domains deep within the KL1–KL2 cleft (Fig. 3c), indicating that their correct apposition is critically important for α-klotho co-receptor activity. These structural observations suggest that the bound Zn²⁺ ion serves as a prosthetic group in α-klotho by minimizing interdomain flexibility and hence promoting co-receptor activity. Consistent with such a role, mutants of membrane-anchored α-klothoTM carrying alanine in place of two, three or all four Zn²⁺ coordinating amino acids (Fig. 3c) showed a reduced ability to support FGF23 signalling (Fig. 4e). Together with our data on the effect of RBA deletion, these results corroborate the biological relevance of the crystallographically deduced mode by which α-klotho implements FGF23–FGFR1c proximity and thus confers high binding affinity.

FGF23 signalling is α-klotho- and HS-dependent

Both FGF23 and FGFR1c have a measurable (albeit weak) binding affinity for HS. Because HS is ubiquitously expressed, we wondered

whether it participates in the apparent α-klotho^{ecto}-mediated FGF23–FGFR dimerization in our cell-based and *in vivo* experiments. We therefore analysed the molecular mass of the ternary complex in the absence and presence of increasing molar equivalents of homogenously sulfated heparin hexasaccharide (HS6). Consistent with our previous observations, in the absence of HS6, the ternary complex migrated as a monomeric species¹⁰ with an apparent molecular mass of 150 kDa, in good agreement with the theoretical value for a 1:1:1 complex (160 kDa) (Fig. 5a). With increasing molar ratios of HS6 to ternary complex, the peak for monomeric ternary complex diminished, while a new peak with a molecular mass of 300 kDa (corresponding to a 2:2:2 FGF23–FGFR1c^{ecto}–α-klotho^{ecto} dimer) appeared and increased in prominence. Excess HS6 beyond a 1:1 molar ratio of HS6 to ternary complex did not lead to any further increase in the amount of dimer complex formed, as judged by the integrated area of the dimer complex peak (Fig. 5a). We conclude that HS is required for the dimerization of 1:1:1 FGF23–FGFR1c^{ecto}–α-klotho^{ecto} complexes, and that at least a 1:1 molar ratio of HS6 to ternary complex is required for complete dimerization of the complex in solution (Fig. 5a). To confirm the dependency of dimerization on HS, we introduced mutations into the HS-binding sites of FGFR1c (K160Q/K163Q, FGFR1c^{ΔHBS}, and K207Q/R209Q, FGFR1c^{ΔHBS'}) and FGF23 (R140A/R143A, FGF23^{ΔHBS}). Neither mutating the HS-binding site in FGFR1c nor mutating that site in FGF23 affected the formation of a monomeric 1:1:1 FGF23–FGFR1c–α-klotho complex in solution, demonstrating that α-klotho-mediated stabilization of the FGF23–FGFR complex is independent of HS. However, ternary complexes containing any of these three mutants failed to dimerize in the presence of HS6 (Fig. 5b).

Reconstitution experiments in the context of BaF3 cells (an FGFR, α-klotho and HS triple-deficient cell line²⁹) showed that both soluble α-klotho^{ecto} and membrane-bound α-klothoTM required HS to support FGF23-mediated FGFR1c activation in a more physiological context (Fig. 5c). We also examined the impact of the HS-binding site mutations in FGFR1c and FGF23 on FGFR1c activation by FGF23 in BaF3 cells (Fig. 5d). In agreement with our solution binding data, activation by FGF23 of HS-binding site mutants of FGFR1c in BaF3 cells was markedly impaired, regardless of whether soluble or membrane-bound α-klotho served as the co-receptor (Fig. 5d). Similarly, the binding site mutant of FGF23 showed a markedly reduced ability to activate FGFR1c (Fig. 5e). These *in vitro* and cell-based analyses unequivocally demonstrate that whereas HS fulfils a dual role in paracrine FGF signalling—enhancing 1:1 FGF–FGFR binding and promoting 2:2 FGF–FGFR dimerization—it shares this task with α-klotho in FGF23 signalling. Thus, α-klotho primarily acts to promote 1:1 FGF23–FGFR1c binding, whereas HS induces the dimerization of the resulting FGF23–FGFR1c–α-klotho complexes.

On the basis of the crystallographically deduced 2:2:2 (Protein Data Bank (PDB) code 1FQ9)⁴ and 2:2:1 (PDB code 1E00)³⁰ paracrine FGF–FGFR–HS dimerization models, two distinct HS-induced 2:2:2 endocrine FGF23–FGFR1c–α-klotho quaternary dimers can be predicted that differ markedly in the composition of the dimer interface (Extended Data Fig. 8). Specifically, in the 2:2:2:1 model, there would be no protein–protein contacts between the two 1:1:1 FGF23–FGFR1c–α-klotho protomers (Extended Data Fig. 8a). By contrast, in the 2:2:2:2 model, FGF23 and FGFR1c from one 1:1:1 FGF23–FGFR1c–α-klotho protomer would interact with the D2 domain of FGFR1c in the adjacent 1:1:1 FGF23–FGFR1c–α-klotho protomer across a two-fold dimer interface (Extended Data Fig. 8b). On the basis of the fundamental differences in the composition of the dimer interface between these two models, we introduced mutations into the secondary-receptor-binding site (SRBS) in FGF23 (M149A/N150A/P151A; FGF23^{ΔSRBS}) and into the corresponding secondary-ligand-binding site (SLBS) in FGFR1c (D2 (I203E, FGFR1c^{ΔSLBS}, and V221D, FGFR1c^{ΔSLBS'}), both of which are unique to the 2:2:2:2 quaternary dimer model. The direct receptor–receptor binding site (RRBS) in FGFR1c D2 (A171D; FGFR1c^{ΔRRBS}), another binding site unique to the 2:2:2:2 model, was also mutated

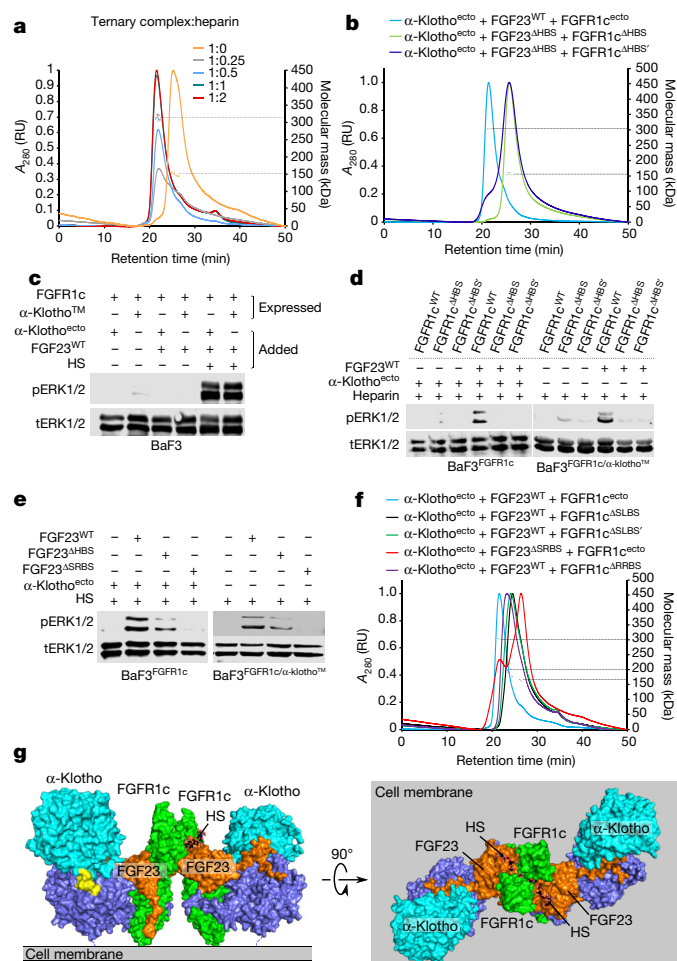


Figure 5 | Heparan sulfate dimerizes two 1:1:1 FGF23-FGFR1c- α -klotho complexes into a symmetric 2:2:2 FGF23-FGFR1c- α -klotho-HS signal transduction unit. a, SEC-MALS analysis of the FGF23-FGFR1c^{ecto}- α -klotho^{ecto} complex in the absence or presence of increasing molar amounts of heparin hexasaccharide (HS6). **b**, SEC-MALS analysis of the FGF23-FGFR1c^{ecto}- α -klotho^{ecto} complexes containing HS-binding site mutations of FGF23 and FGFR1c. **c–e**, Representative immunoblots of phosphorylated ERK (top) and total ERK (bottom; sample loading controls) in total BaF3 cell lysates ($n = 3$ independent experiments for each panel). **c**, Analysis of HS dependency of FGF23 signalling. **d**, **e**, Analysis of mutations in the HS-binding site of FGFR1c (**d**) and in the HS-binding site or secondary receptor-binding site of FGF23 (**e**). **f**, SEC-MALS analysis of FGF23-FGFR1c^{ecto}- α -klotho^{ecto} complexes containing a secondary receptor-binding site mutation in FGF23, a secondary ligand-binding site mutation in FGFR1c, or a direct receptor-receptor-binding site mutation in FGFR1c. In **b** and **f**, wild-type ternary complex served as controls. **g**, Molecular surface of a 2:2:2 FGF23-FGFR1c- α -klotho-HS dimer in two orientations related by a 90° rotation around the horizontal axis: a side-view looking parallel to the plane of a cell membrane (left) and a bird's-eye view looking down onto the plane of a cell membrane (right). HS molecules are shown as black sticks.

(Extended Data Fig. 8b). Although all of these FGF23 and FGFR1c mutants were able to form ternary complexes with α -klotho^{ecto}, the ternary complexes containing any of the mutated proteins were impaired in their ability to dimerize in the presence of HS6 in solution (Fig. 5f). Moreover, the FGF23^{ΔSRBS} mutant showed a markedly diminished ability to activate FGFR1c in BaF3 cells (Fig. 5e). The loss-of-function effects of these mutations are consistent with a 2:2:2 quaternary dimer model (Extended Data Fig. 8b). Hence, we envision that HS engages the HS-binding sites of FGFR1c and FGF23 in two stabilized 1:1:1 FGF23-FGFR1c- α -klotho ternary complexes to promote the formation of a two-fold symmetric 2:2:2

FGF23-FGFR1c- α -klotho-HS dimer (Fig. 5g). In doing so, HS enhances reciprocal interactions of FGFR1c D2 and FGF23 from one ternary complex with FGFR1c D2 in the other ternary complex, thereby buttressing the dimer (Extended Data Fig. 8b). This replicates the role that HS has in paracrine FGF signalling⁴. In contrast to HS, α -klotho molecules do not directly participate in the dimer interface (Fig. 5g), but rather indirectly support HS-induced dimerization by enhancing 1:1 FGF23-FGFR1c binding affinity. Hence, FGF23 seems to strike a fine balance between losing a large amount of HS-binding affinity to enable its endocrine mode of action and retaining sufficient HS-binding affinity to allow HS-mediated dimerization of two 1:1:1 FGF23-FGFR1c- α -klotho complexes. These considerations do not formally exclude the possibility that 2:2:2:2 and 2:2:2:1 quaternary dimers might co-exist as a higher order cluster on the cell surface, as has been proposed previously for paracrine 2:2:2 and 2:2:1 FGF-FGFR1-HS dimers³¹.

FGF19 and FGF21, the other two endocrine FGFs, both require β -klotho as an obligate co-receptor to bind and activate cognate FGFRs^{32,33} so as to mediate effects that regulate, for example, metabolic pathways involved in bile acid biosynthesis or fatty acid oxidation^{34,35}. On the basis of the structural analysis and supporting cell-based data shown in Extended Data Figs 9 and 10, we propose that β -klotho, similar to α -klotho, functions as a non-enzymatic molecular scaffold to promote signalling by these two FGF hormones.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 June; accepted 13 December 2017.

Published online 17 January 2018.

- Shimada, T. *et al.* Targeted ablation of Fgf23 demonstrates an essential physiological role of FGF23 in phosphate and vitamin D metabolism. *J. Clin. Invest.* **113**, 561–568 (2004).
- Gattineni, J. *et al.* FGF23 decreases renal NaPi-2a and NaPi-2c expression and induces hypophosphatemia *in vivo* predominantly via FGF receptor 1. *Am. J. Physiol. Renal Physiol.* **297**, F282–F291 (2009).
- Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
- Schlessinger, J. *et al.* Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol. Cell* **6**, 743–750 (2000).
- Mohammadi, M., Olsen, S. K. & Ibrahimi, O. A. Structural basis for fibroblast growth factor receptor activation. *Cytokine Growth Factor Rev.* **16**, 107–137 (2005).
- Goetz, R. & Mohammadi, M. Exploring mechanisms of FGF signalling through the lens of structural biology. *Nat. Rev. Mol. Cell Biol.* **14**, 166–180 (2013).
- Kuro-o, M. *et al.* Mutation of the mouse klotho gene leads to a syndrome resembling ageing. *Nature* **390**, 45–51 (1997).
- Henrissat, B. & Davies, G. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* **7**, 637–644 (1997).
- Goetz, R. *et al.* Klotho coreceptors inhibit signaling by paracrine fibroblast growth factor 8 subfamily ligands. *Mol. Cell Biol.* **32**, 1944–1954 (2012).
- Goetz, R. *et al.* Isolated C-terminal tail of FGF23 alleviates hypophosphatemia by inhibiting FGF23-FGFR-Klotho complex formation. *Proc. Natl Acad. Sci. USA* **107**, 407–412 (2010).
- Urakawa, I. *et al.* Klotho converts canonical FGF receptor into a specific receptor for FGF23. *Nature* **444**, 770–774 (2006).
- Kurosu, H. *et al.* Regulation of fibroblast growth factor-23 signaling by klotho. *J. Biol. Chem.* **281**, 6120–6123 (2006).
- Li, S. A. *et al.* Immunohistochemical localization of Klotho protein in brain, kidney, and reproductive organs of mice. *Cell Struct. Funct.* **29**, 91–99 (2004).
- van Loon, E. P. *et al.* Shedding of klotho by ADAMs in the kidney. *Am. J. Physiol. Renal Physiol.* **309**, F359–F368 (2015).
- Lindberg, K. *et al.* The kidney is the principal organ mediating klotho effects. *J. Am. Soc. Nephrol.* **25**, 2169–2175 (2014).
- Chen, C. D., Podvin, S., Gillespie, E., Leeman, S. E. & Abraham, C. R. Insulin stimulates the cleavage and release of the extracellular domain of Klotho by ADAM10 and ADAM17. *Proc. Natl Acad. Sci. USA* **104**, 19796–19801 (2007).
- Imura, A. *et al.* Secreted Klotho protein in sera and CSF: implication for post-translational cleavage in release of Klotho protein from cell membrane. *FEBS Lett.* **565**, 143–147 (2004).
- Matsumura, Y. *et al.* Identification of the human klotho gene and its two transcripts encoding membrane and secreted klotho protein. *Biochem. Biophys. Res. Commun.* **242**, 626–630 (1998).
- Shiraki-Iida, T. *et al.* Structure of the mouse klotho gene and its two transcripts encoding membrane and secreted protein. *FEBS Lett.* **424**, 6–10 (1998).

20. Kurosu, H. *et al.* Suppression of aging in mice by the hormone Klotho. *Science* **309**, 1829–1833 (2005).
21. Hu, M. C., Shiizaki, K., Kuro-o, M. & Moe, O. W. Fibroblast growth factor 23 and Klotho: physiology and pathophysiology of an endocrine network of mineral metabolism. *Annu. Rev. Physiol.* **75**, 503–533 (2013).
22. Chang, Q. *et al.* The β -glucuronidase klotho hydrolyzes and activates the TRPV5 channel. *Science* **310**, 490–493 (2005).
23. Cha, S. K. *et al.* Removal of sialic acid involving Klotho causes cell-surface retention of TRPV5 channel via binding to galectin-1. *Proc. Natl Acad. Sci. USA* **105**, 9805–9810 (2008).
24. Hu, M. C. *et al.* Klotho: a novel phosphaturic substance acting as an autocrine enzyme in the renal proximal tubule. *FASEB J.* **24**, 3438–3450 (2010).
25. Imura, A. *et al.* α -Klotho as a regulator of calcium homeostasis. *Science* **316**, 1615–1618 (2007).
26. Hayashi, Y. *et al.* Klotho-related protein is a novel cytosolic neutral beta-glycosylceramidase. *J. Biol. Chem.* **282**, 30889–30900 (2007).
27. Andrukhova, O. *et al.* Klotho lacks an FGF23-independent role in mineral homeostasis. *J. Bone Miner. Res.* **32**, 2049–2061 (2017).
28. Wu, X. *et al.* C-terminal tail of FGF19 determines its specificity toward Klotho co-receptors. *J. Biol. Chem.* **283**, 33304–33309 (2008).
29. Ornitz, D. M. *et al.* Heparin is required for cell-free binding of basic fibroblast growth factor to a soluble receptor and for mitogenesis in whole cells. *Mol. Cell. Biol.* **12**, 240–247 (1992).
30. Pellegrini, L., Burke, D. F., von Delft, F., Mulloy, B. & Blundell, T. L. Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature* **407**, 1029–1034 (2000).
31. Harmer, N. J. *et al.* Towards a resolution of the stoichiometry of the fibroblast growth factor (FGF)–FGF receptor–heparin complex. *J. Mol. Biol.* **339**, 821–834 (2004).
32. Ogawa, Y. *et al.* β Klotho is required for metabolic activity of fibroblast growth factor 21. *Proc. Natl Acad. Sci. USA* **104**, 7432–7437 (2007).
33. Kurosu, H. *et al.* Tissue-specific expression of β Klotho and fibroblast growth factor (FGF) receptor isoforms determines metabolic activity of FGF19 and FGF21. *J. Biol. Chem.* **282**, 26687–26695 (2007).
34. Holt, J. A. *et al.* Definition of a novel growth factor-dependent signal cascade for the suppression of bile acid biosynthesis. *Genes Dev.* **17**, 1581–1591 (2003).
35. Potthoff, M. J. *et al.* FGF21 induces PGC-1 α and regulates carbohydrate and fatty acid metabolism during the adaptive starvation response. *Proc. Natl Acad. Sci. USA* **106**, 10853–10858 (2009).
36. Pitteloud, N. *et al.* Digenic mutations account for variable phenotypes in idiopathic hypogonadotropic hypogonadism. *J. Clin. Invest.* **117**, 457–463 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank N. J. Cowan for critically reading and editing the manuscript, and C.-S. Huang for help with diffraction data processing with XDS. This work was primarily supported by NIH grant R01 DE13686 (to M.M.). Support was also provided by National Key R&D Program of China (#2017YFA0506000 to X.L.). Funding for mouse studies was provided by R01 DK092461, P30 DK079328 (to O.W.M.), and R01 DK091392 (to M.C.H.). Beamlines at the Northeastern Collaborative Access Team (NE-CAT) facility at the Advanced Photon Source of Argonne National Laboratory are primarily funded by NIH NIGMS and member institutions.

Author Contributions G.C. purified and crystallized the ternary complex, analysed the crystal structure, generated SEC–MALS data (Figs 4a, 5a, b, f), cell-based data (Fig. 4), enzyme and thermostability assay data (Fig. 2c), and participated in the design of experiments and the writing/revising of the manuscript. Y.L. helped with data collection and analysis of the crystal structure, generated cell-based data (Fig. 5), and participated in manuscript revision. R.G. established expression and purification protocols for the ternary complex, performed ternary complex characterization, analysed mouse data, and participated in editing and revising the manuscript. L.F. generated expression constructs for FGF23, FGFR1^{ecto}, α -klotho^{ecto} and their structure-based mutated forms, and helped with ternary complex purification. S.J. assisted with diffraction data collection and performed excitation/emission scanning of the FGF23–FGFR1^{ecto}– α -klotho^{ecto} crystal (Extended Data Fig. 2c). M.-C.H. and O.W.M. generated the mouse data (Extended Data Figs 1c, d and 7a, b). G.L. and X.L. (mentors of G.C. and L.F.) participated in manuscript revision. M.M. developed and directed the project, solved, refined, analysed and interpreted the crystal structure of the ternary complex, and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.M. (moosa.mohammadi@nyumc.org) or X.L. (xiaokunli@wzmu.edu.cn).

Reviewer Information *Nature* thanks M. Kuro-o and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and, except for the data shown in Extended Data Fig. 7a, b, investigators were not blinded to allocation during experiments and outcome assessment.

DNA expression constructs. cDNA fragments encoding full-length human α -klotho, β -klotho and FGFR1c were amplified by PCR and subcloned into the lentiviral transfer plasmids pEF1 α -IRES-hygro (α -/ β -klotho) or pEF1 α -IRES-Neo (FGFR1c) using a ligation-independent In-Fusion HD cloning kit (639648, Clontech Laboratories). PCR primers for FGFR1 'c' isotype were designed using NEBaseChanger software version 1.2.6 (New England Biolabs) and primers for *KL* and *KLB* (encoding α -klotho and β -klotho, respectively) were designed using the primer design tool for the In-Fusion HD cloning kit (Clontech Laboratories). A cDNA fragment encoding the entire extracellular domain of human α -klotho (residues Met1 to Ser981; α -klotho^{ecto}) was subcloned into the mammalian expression plasmid pEF1 α /myc-His A. DNA fragments for the mature form (that is, without the signal sequence) of human FGF23 (residues Tyr25 to Ile251), human FGF21 (residues His29 to Ser209), and the extracellular D2–D3 region of human FGFR1c (residues Asp142 to Arg365; FGFR1c^{ecto}), which is both necessary and sufficient for FGF binding, were amplified by PCR and ligated into the cloning sites of the bacterial expression plasmids pET-30a and pET-28a, respectively. Single/multiple site mutations, loop deletions and truncations were introduced into expression constructs encoding the wild-type proteins using a Q5 Site-Directed Mutagenesis Kit (E0554S, New England Biolabs). The integrity of each expression construct was confirmed by restriction enzyme digestion and DNA sequencing. Information on the constructs is provided in the Supplementary Tables 1 and 2.

Recombinant protein expression and purification. *N*-acetylglucosaminyltransferase I (GnTI) deficient HEK293S cells (CRL-3022, American Type Culture Collection (ATCC)) were transfected by calcium phosphate co-precipitation with the expression construct encoding α -klotho^{ecto}. G418-resistant colonies were selected for α -klotho^{ecto} expression using 0.5 mg ml⁻¹ G418 (6483, KSE Scientific). The clone with the highest expression level was propagated in DME/F12 medium (SH30023.02, HyClone) supplemented with 10% fetal bovine serum (FBS) (35-010-CV, CORNING), 100 U ml⁻¹ penicillin plus 100 μ g ml⁻¹ streptomycin (15140-122, Gibco), and 0.5 mg ml⁻¹ G418. For protein production, 1×10^6 cells were seeded in 25 cm cell culture dishes in 20 ml DME/F12 medium containing 10% FBS and grown for 24 h. Thereafter, the medium was replaced with 25 ml DME/F12 medium containing 1% FBS. Three days later, secreted α -klotho^{ecto} from two litres of conditioned medium was captured on a 5 ml heparin affinity HiTrap column (GE Healthcare) and eluted with a 100 ml linear NaCl gradient (0–1.0 M). Column fractions containing α -klotho^{ecto} were pooled and diluted tenfold with 25 mM Tris pH 8.0 buffer, and the diluted protein sample was loaded onto an anion exchange column (SOURCE Q, GE Healthcare) and eluted with a 280 ml linear NaCl gradient (0–0.4 M). As a final purification step, SOURCE Q fractions containing α -klotho^{ecto} were concentrated and applied to a Superdex 200 column (GE Healthcare). α -Klotho^{ecto} protein was eluted isocratically in 25 mM HEPES pH 7.5 buffer containing 500 mM NaCl and 100 mM (NH₄)₂SO₄. A mutant of α -klotho^{ecto} lacking the receptor binding arm (α -klotho^{ecto}/ Δ RBA) was expressed and purified similarly as the wild-type counterpart.

Human wild-type FGF23 and its mutants were expressed in *Escherichia coli* BL21 DE3 cells. Inclusion bodies enriched in misfolded insoluble FGF23 protein were dissolved in 6 M guanidinium hydrochloride and FGF23 proteins were refolded by dialysis for 2 days at 4 °C against buffer A (25 mM HEPES pH 7.5, 150 mM NaCl, 7.5% glycerol) followed by buffer B (25 mM HEPES pH 7.5, 100 mM NaCl, 5% glycerol). Correctly folded FGF23 proteins were captured on a 5 ml heparin affinity HiTrap column (GE Healthcare) and eluted with a 100 ml linear NaCl gradient (0–2.0 M). Final purification of FGF23 proteins was achieved by cation exchange chromatography (SOURCE S, GE Healthcare) with a 280 ml linear NaCl gradient (0–0.4 M). Human FGFR1c^{ecto} and its mutants were also expressed as inclusion bodies in *E. coli* BL21 DE3 and refolded *in vitro* by slow dialysis at 4 °C against the following buffers: buffer A (25 mM Tris pH 8.2, 150 mM NaCl, 7.5% glycerol), buffer B (25 mM Tris pH 8.2, 100 mM NaCl, 5% glycerol), and buffer C (25 mM Tris pH 8.2, 50 mM NaCl, 5% glycerol); dialysis against each buffer was for minimally 12 h. Properly folded FGFR1c proteins were purified by heparin affinity chromatography followed by size-exclusion chromatography as described above. All column chromatography was performed at 4 °C on an AKTA pure 25 I system (GE Healthcare).

Crystallization and X-ray crystal structure determination. To facilitate crystallization of the FGF23–FGFR1c^{ecto}– α -klotho^{ecto} complex, we used a proteolytically and structurally more stable FGF23 protein variant, which lacked 46 residues from the FGF23 C-terminus (Cys206 to Ile251) and carried Arg-to-Gln mutations at positions 176 and 179 of the 176-Arg-His-Thr-Arg-179 proteolytic cleavage motif in FGF23. The Arg-to-Gln mutations occur naturally in patients

with autosomal dominant hypophosphatemic rickets (ADHR)³⁷, and deletion of C-terminal residues Cys206 to Ile251 has no effect on the phosphaturic activity of FGF23 in mice or its signalling potential in α -klothoTM-expressing cultured cells¹⁰. Thus, the first 26 amino acids (Ser180 to Ser205) of the 72-amino-acid-long C-terminal tail of FGF23, defined as the region past the 176-Arg-His-Thr-Arg-179 proteolytic cleavage site, comprise the minimal region of the FGF23 C-terminal tail for binding the FGFR1c^{ecto}– α -klotho^{ecto} complex¹⁰. To prepare the FGF23–FGFR1c^{ecto}– α -klotho^{ecto} complex, its purified components were mixed at a molar ratio of 1.2:1.2:1 and spin-concentrated using an Amicon Ultra-15 concentrator (UFC901024, Merck Millipore). The concentrated sample was applied to a Superdex 200 column (GE Healthcare) and eluted isocratically in 25 mM HEPES pH 7.5 buffer containing 500 mM NaCl and 100 mM (NH₄)₂SO₄. Column peak fractions were analysed by SDS–PAGE and peak fractions containing the ternary complex were concentrated to 7 mg ml⁻¹. Concentrated ternary complex was screened for crystallization by sitting drop vapour diffusion. A range of commercially available crystallization screen kits was used: Protein Complex Suite (130715), Classics Suite (130701), Classics II Suite (130723), and Classics Lite Suite (130702) from Qiagen; Crystal Screen (HR2-110), Crystal Screen 2 (HR2-112), Crystal Screen Lite (HR2-128), PEG/Ion Screen (HR2-126), and PEGRx1 (HR2-082) from Hampton Research; and PEG Grid Screening Kit (36436) and Crystallization Cryo Kit (75403) from Sigma-Aldrich. Drops consisting of 100 nl reservoir solution and 100 nl protein complex solution were equilibrated against 100 μ l well volume set up in 96-well plates (Fisher Scientific) using a Mosquito crystallization robot (TTP Labtech). Plates were stored at 18 °C and automatically imaged by Rock Imager 1000 (Formulatrix). Image data were collected and managed using Rock Maker software version 3.1.4.0 (Formulatrix). One crystal hit was obtained after 7 days of plate incubation at 18 °C and one crystallization condition from the Protein Complex Suite (130715, Qiagen) was chosen for optimization using the Additive Screen (HR2-428) from Hampton Research. Crystals were confirmed as protein crystals by UV imaging using Rock Imager 1000 (Formulatrix). Crystal growth in optimized conditions was scaled up in 24-well VDXm plates (Hampton Research) where crystals were grown by hanging drop vapour diffusion. Larger crystals (80 \times 76 \times 35 μ m) were obtained within 28 days by mixing 1 μ l of protein complex and 1 μ l of crystallization solution. Some of those crystals were dissolved in Lämmli sample buffer after thorough rinsing, and analysed by SDS–PAGE and staining with Coomassie blue to confirm the presence of all three proteins in the ternary complex.

Crystals of ternary complex were briefly soaked in cryo-protective solution consisting of mother liquor supplemented with 25% (w/v) glycerol. These were then mounted on CryoLoops (Hampton Research) and flash-frozen in liquid nitrogen. Crystal screening for X-ray diffraction and diffraction data collection were performed at 100 K on one of the NE-CAT beam lines at the Advanced Photon Source synchrotron of Argonne National Laboratory. X-ray images were recorded with an ADSC Quantum 315 CCD detector with primary oscillations at 100 K, a wavelength of 0.97918 Å, and a crystal-to-detector distance of 420 mm. Crystals of the ternary complex belong to the monoclinic space group C2, and contain one ternary complex molecule in the asymmetric unit. X-ray diffraction data sets were collected to 3.0 Å from native protein crystals, integrated, and scaled using XDS³⁸ and SCALA³⁹ from the CCP4 software suite⁴⁰.

A clear molecular replacement solution was found for both KL domains using the Phaser module of PHENIX⁴¹ and homology models of KL1 and KL2, which were built with Rosetta software available through the ROSETTA Protein Structure Prediction Server (<http://rosetta.bakerlab.org>). However, the FGF23–FGFR1c component of the ternary complex could not be found even after fixing the coordinates of the partial solution found for the KL domains. Through careful inspection of the crystal lattice and the $F_o - F_c$ difference and $2F_o - F_c$ composite maps generated using the partial model, we succeeded in manually placing an FGF23–FGFR1c D2 portion of the FGF23–FGFR1c complex. This was created using the experimental crystal structures of SOS-bound FGF23⁴² (PDB code 2P39) and the FGF2-bound FGFR1c ectodomain⁴³ (PDB code 1CVS). After a few rounds of refinements, FGFR1c D3 could also be placed manually. Iterative rounds of model building and refinement were carried out using Coot⁴⁴ and the Phenix. Refine module of PHENIX⁴¹.

The structure has been refined to 3.0 Å resolution with working and free *R*-factors of 23.46 and 28.26%, respectively, and good Ramachandran plot statistics. X-ray diffraction data collection and structure refinement statistics are summarized in Extended Data Table 1. The final model comprises residues Glu34 to His977 of human α -klotho^{ecto}, residues Met149 to Ala361 of human FGFR1c^{ecto} and residues Tyr25 to Thr200 of human FGF23. Owing to insufficient electron density, the following residues of the ternary complex could not be built: 1) Leu98 to Ser115 (β 1 α 1 loop) of α -klotho^{ecto} KL1, 2) Glu957 to Glu960 (an ADAM protease cleavage site) at the junction between the rigid core of α -klotho^{ecto} KL2 and the flexible extracellular juxtamembrane linker that connects KL2 to the transmem-

brane helix of α -klotho, 3) the last four residues of the extracellular juxtamembrane linker (Thr978 to Ser981) of α -klotho^{ecto}, 4) the last five C-terminal residues of FGF23 (Pro201 to Ser205), 5) Asp142 to Arg148 N-terminal to the D2 domain of FGFR1c^{ecto}, and 6) Leu362 to Arg365 C-terminal to the D3 domain of FGFR1c^{ecto}. Ordering of the first six N-terminal residues of FGF23 (Tyr25 to Pro30) is influenced by crystal lattice contacts.

SEC–MALS. The SEC–MALS instrument setup consisted of a Waters Breeze 2 HPLC system (Waters), a miniDAWN-TREOS 18-angle static light scattering detector with built-in 658.0-nm wavelength laser (Wyatt Technology Corp.), and an Optilab rEX refractive index detector (Wyatt Technology Corp.). A Superdex 200 10/300 GL column (GE Healthcare) was placed in-line between the HPLC pump (Waters 1525) and the HPLC UV (Waters 2998 Photodiode Array), laser light scattering, and refractive index detectors. Light scattering and refractive index detectors were calibrated following the manufacturer's guidelines. The refractive index increment (dn/dc), in which n is the refractive index and c is the concentration of the mixture of DDM and CHS in 20 mM Tris-HCl pH 8.0 buffer containing 300 mM NaCl, was determined offline using an Optilab T-REX refractive index detector. Monomeric bovine serum albumin (23210, Thermo Scientific) was used as part of routine data quality control.

At least 60 ml of 25 mM HEPES pH 7.5 buffer containing 150 mM NaCl were passed through the system at a flow rate of 0.5 ml min⁻¹ to equilibrate the Superdex 200 10/300 GL column and establish stable baselines for light scattering and refractive index detectors. Purified α -klotho^{ecto}, FGFR1c^{ecto} (wild type or mutant), and FGF23 (wild type or mutant) proteins were mixed at a molar ratio of 1:1:1 and concentrated to 12.5 μ M. Protein samples (50 μ l) with a molar equivalent of a heparin hexasaccharide (HO06, Iduron) were injected onto the gel filtration column, and the column eluent was continuously monitored for 280 nm absorbance, laser light scattering, and refractive index. In a separate set of experiments, 50 μ l of 1:1:1 FGF23–FGFR1c^{ecto}– α -klotho^{ecto} ternary complex at 12.5 μ M concentration was mixed with heparin hexasaccharide at molar ratios of 1:0.25, 1:0.5, 1:1 or 1:2, and the mixtures were injected onto the gel filtration column. As a control, 50 μ l of ternary complex without added heparin hexasaccharide were run on the column. In yet another set of experiments, α -klotho^{ecto} (wild type or mutant) and FGFR1c^{ecto} were mixed at a molar ratio of 1:1, and 50 μ l of concentrated protein mixtures were injected onto the gel filtration column. 50 μ l of concentrated α -klotho^{ecto} (wild type or mutant) alone were run as a control in these experiments. The analyses were performed at ambient temperature. Data were collected every second at a flow rate of 0.5 ml min⁻¹. Laser light scattering intensity and eluent refractive index (concentration) data were adjusted manually for the volume delay of UV absorbance at 280 nm, and were processed using ASTRA software (Wyatt Technology Corp.). A protein refractive index increment (dn/dc value) of 0.185 ml g⁻¹ was used for molecular mass calculations.

Cell line culture and stimulation and analysis of protein phosphorylation. HEK293 cells (a gift from A. Mansukhani, identified by morphology check under microscope, mycoplasma negative in DAPI) were maintained in DMEM medium (10-017-CV, CORNING) supplemented with 10% FBS, 100 U ml⁻¹ of penicillin and 100 μ g ml⁻¹ streptomycin. HEK293 cells naturally express multiple FGFR isoforms including FGFR1c, FGFR3c and FGFR4, but lack α -klotho or β -klotho co-receptors. BaF3 cells (a gift from S. Byron, identified by morphology check under microscope, mycoplasma negative in DAPI), an IL-3-dependent haematopoietic pro B cell line, were cultured in RPMI 1640 medium (10-040-CV, CORNING) supplemented with 10% FBS, 100 U ml⁻¹ of penicillin, 100 μ g ml⁻¹ streptomycin and 5 ng ml⁻¹ mouse IL-3 (#GFM1, Cell Guidance Systems). BaF3 cells do not express FGFRs, α -/ β -klotho co-receptors, or HS cofactors, and hence are naturally non-responsive to FGFs. However, via controlled ectopic expression of FGFRs and klotho co-receptors and exogenous supplementation with soluble HS, these cells can be forced to respond to FGF stimulation. As such, the BaF3 cell line has served as a powerful tool for reconstituting FGF–FGFR cell surface signal transduction complexes to dissect the molecular mechanisms of paracrine and endocrine FGF signalling^{29,45,46}.

Stable or transient expression of full-length (transmembrane) human α -klotho, β -klotho, FGFR1c, and mutants of these proteins in HEK293 or BaF3 cells was achieved using lentiviral vectors. To generate lentiviral expression vectors, HEK293 cells were seeded at a density of about 8×10^5 in 10 cm cell culture dishes and co-transfected by calcium phosphate co-precipitation with 8 μ g of lentiviral transfer plasmid encoding wild-type or mutant α -klotho, β -klotho or FGFR1c, 1.6 μ g of pMD2.G envelope plasmid, and 2.5 μ g of psPAX2 packaging plasmid. Fresh medium was added to the cells for a 3-day period after transfection. Cell culture supernatant containing recombinant lentivirus particles was collected and used to infect 2×10^5 HEK293 or BaF3 cells in the presence of polybrene (5 μ g ml⁻¹; 134220, Santa Cruz Biotechnology). Stable transfectants were selected using hygromycin (1 mg ml⁻¹, ant-hg-1, InvivoGen) or G418 (0.5 mg ml⁻¹, 6483, KSE Scientific). For transient protein expression, 2×10^5 HEK293 cells were plated

in 6-well cell culture dishes and on the following day, the cells were infected with recombinant lentivirus in the presence of polybrene (16 μ g).

For cell stimulation studies, unmodified and stably transfected HEK293 cells were seeded in 6-well cell culture plates at a density of 4×10^5 cells per well and maintained for 24 h in cell culture medium without FBS. In the case of transiently transfected HEK293 cells, medium containing lentivirus particles was removed from the cells after incubation for approximately 12 h, and the cells were also serum-starved for 24 h. Stably transfected BaF3 cells were seeded in 10 cm cell culture dishes at a density of 6×10^6 cells and serum-starved for 6 h. Unmodified HEK293 cells were stimulated for 10 min with wild-type or mutant FGF23 both in the presence and absence of wild-type or mutant α -klotho^{ecto}. HEK293 cells stably or transiently expressing wild-type α -klotho^{ecto} or its mutants were stimulated with wild-type or mutant FGF23 alone. In one set of experiments, HEK293 cells expressing wild-type α -klotho^{ecto} were pretreated with α -klotho^{ecto} for 10 min before stimulation with wild-type FGF23. BaF3 cells expressing wild-type or mutant FGFR1c were stimulated with wild-type or mutant FGF23 in the presence or absence of α -klotho^{ecto} and heparin. BaF3 cells co-expressing wild-type α -klotho^{ecto} and wild-type or mutant FGFR1c were stimulated with wild-type or mutant FGF23 in the presence of heparin. BaF3 cells co-expressing wild-type FGFR1c and wild-type or mutant β -klotho^{ecto} were stimulated with wild-type FGF21 in the presence or absence of heparin.

After stimulation, cells were lysed, and lysate samples containing approximately 30 μ g total cellular protein were electrophoresed on 12% SDS–PAGE and electrotransferred onto a nitrocellulose membrane. The membrane was blocked for 1 h at ambient temperature in Tris-buffered saline pH 7.6 containing 0.05% Tween-20 and 5% BSA (BP1600-100, Fisher BioReagents). Rabbit monoclonal antibodies to phosphorylated ERK1/2 (4370, Cell Signaling Technology) and total (phosphorylated and unphosphorylated) ERK1/2 (4695, Cell Signaling Technology) were diluted 1:2,000 and 1:1,000, respectively, in blocking buffer. After overnight incubation at 4 °C with one of these diluted antibodies, the blot was washed with Tris-buffered saline pH 7.6 containing 0.05% Tween-20, and then incubated at ambient temperature for 30 min with 1:10,000-diluted IRDye secondary antibody (926-32211 (goat anti-rabbit), LI-COR). After another round of washing with Tris-buffered saline pH 7.6 containing 0.05% Tween-20, the blot was imaged on an Odyssey Fc Dual-mode Imaging System (LI-COR).

α -Klotho treatment of mice and serum, urinary phosphate analysis. Mice of the strain 129/Sv (Charles River Laboratories) were housed in a room with a temperature of 22 ± 1 °C and a 12 h:12 h light/dark cycle, and had ad libitum access to tap water and Teklad global 16% rodent diet (Envigo). Ten female and ten male 6-week-old mice of each gender were assigned to receive either recombinant α -klotho^{ecto} protein diluted in isotonic saline (0.1 mg kg⁻¹ body weight) or protein diluent only (buffer control). Mice were placed in metabolic cages for a one-day acclimation, and returned to the cages for 24-h urine collection after intraperitoneal injection of α -klotho^{ecto} protein or buffer control. After urine collection, mice were placed under isoflurane anaesthesia, and blood was drawn from the retro-orbital sinus and transferred into tubes containing a few drops of sterile solution of heparin (Sagent Pharmaceuticals). After centrifugation at 3,000g at 4 °C for 5 min, supernatant plasma was taken out of the tubes and stored at –80 °C. Blood and urine samples were also collected before injection of α -klotho^{ecto} or buffer control. Phosphate and creatinine concentrations in plasma and urine were measured using a Vitros Chemistry Analyzer (Ortho-Clinical Diagnosis) and a P/ACE MDQ Capillary Electrophoresis System equipped with a photodiode detector (Beckman-Coulter), respectively. The Mouse Metabolic Phenotyping Core Facility at UT Southwestern Medical Center carried out the measurements of these analytes.

In a separate set of experiments, 10- to 12-week-old mice were given an intraperitoneal injection of wild-type α -klotho^{ecto} (0.1 mg kg⁻¹ body weight), RBA deletion mutant, α -klotho^{ecto/ΔRBA} (0.1 mg kg⁻¹ body weight), or protein diluent only (three female and three male mice per group), and blood and urine samples were collected for measurement of phosphate and creatinine as described above. In yet another set of experiments, 10- to 12-week-old mice were injected intraperitoneally with 0.1 mg kg⁻¹ body weight of wild-type α -klotho^{ecto} (two female and one male mice), mutant α -klotho^{ecto/ΔRBA} (two female and two male mice), or protein diluent only (two female and one male mice), and kidneys were obtained from the mice under isoflurane anaesthesia four hours after the injection. Total RNA was extracted from the kidneys using RNeasy kit (Qiagen), and *Egr1* mRNA levels were quantified by quantitative PCR (qPCR) with cyclophilin (also known as *Ppia*) as a control. Template cDNA for the PCR was generated using SuperScript III First Strand Synthesis System (Invitrogen) and oligo-(dT) primers. PCR primers for *Egr1* were 5'-GAGGAGATGATGCTGCTGAG-3' and 5'-TGCTGCTGCTGCTATTACC-3'. PCR primers for cyclophilin were 5'-GTCTCTTTTCGCCGCTTGTCT-3' and 5'-TCTGCTGTCTTTGGAAGCTTTGTCTG-3'. qPCR was performed in

triplicate for each kidney RNA sample. Except for *Egr1* expression analysis, data were analysed by paired Student's *t*-test. All studies in mice were approved by the Institutional Animal Care and Use Committee at the University of Texas Southwestern Medical Center and conducted following the National Institutes of Health Guide for the Care and Use of Laboratory Animals.

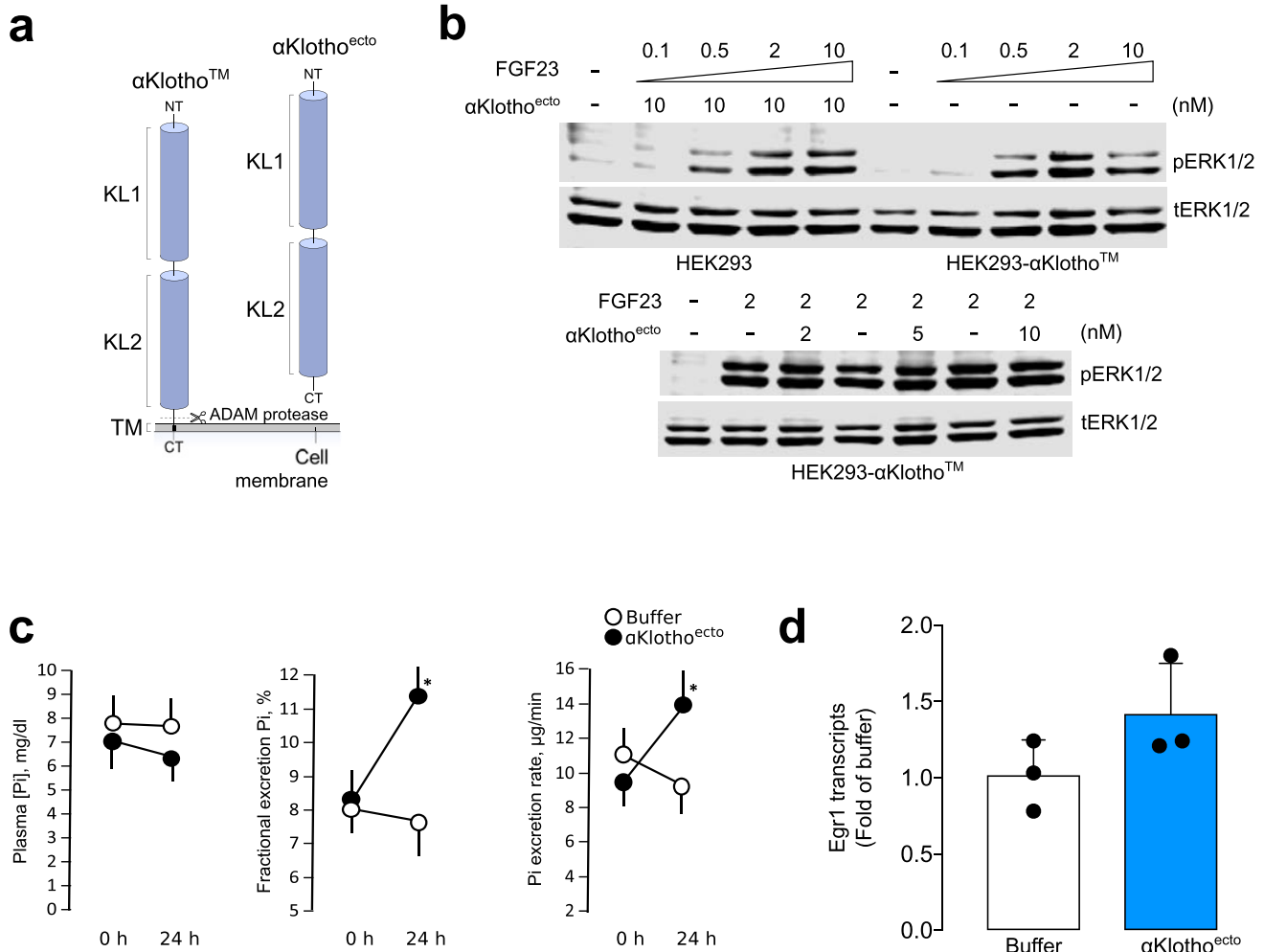
Enzymatic assay. To examine α -klotho^{ecto} for glycoside-hydrolase activity, 4-methylumbelliferyl- β -D-xylopyranoside (M7008, Sigma-Aldrich), 4-methylumbelliferyl- β -D-glucuronide (474427, Sigma-Aldrich) and 4-methylumbelliferyl- α -D-N-acetylneuraminic acid (69587, Sigma-Aldrich) were selected as substrates and commercially available recombinant neuraminidase (#10269611001, Roche Diagnostics GmbH) and β -Glucuronidase (#G0251, Sigma-Aldrich) were used as positive controls. 20 μ g of α -klotho^{ecto} or the control enzymes were added into reaction buffer (0.1 M sodium citrate buffer, pH 5.6, 0.05 M NaCl, 0.01% Tween 20) containing 0.5 mM substrate at a final volume of 100 μ l, and the reaction mixtures were incubated at 37 °C for 2 h. Enzymatic activity was assessed by quantifying fluorescence intensity of released 4-methylumbelliferone at an excitation wavelength of 360 nm and an emission wavelength of 450 nm using a FlexStation 3 Multi-Mode Microplate Reader (Molecular Devices).

Fluorescence dye-based thermal shift assay. SYPRO Orange dye (S6650, ThermoFisher Scientific) was used as the fluorescent probe. 15 μ l of 20 μ M solutions of protein samples (wild-type and mutated forms of FGF23; α -klotho^{ecto} or α -klotho^{ecto/ Δ RBA} alone; 1:1 mixtures of α -klotho^{ecto} or α -klotho^{ecto/ Δ RBA} with FGF23 C-terminal tail peptide) were mixed with 5 μ l of working dye solution (1:25 dilution) in duplicate in PCR strips. A temperature gradient from 4 °C to 100 °C, at 1 °C per min increment was carried out with a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Fluorescence was recorded as a function of temperature in real time. The melting temperature (T_m) was calculated with StepOne software v2.2 as the maximum of the derivative of the resulting SYPRO Orange fluorescence curves.

Statistics and reproducibility. Glycoside-hydrolase activity of α -klotho^{ecto}, neuraminidase and β -glucuronidase was measured in triplicate; one triplicate representative of three independent experiments is shown in Fig. 2c. Each set of immunoblot experiments (data shown in Figs 4b–e, 5c–e and Extended Data Figs 1b, 4c, 7e and 10b, c) was independently repeated three times. Renal mRNA levels of mouse *Egr1* and cyclophilin were each measured in triplicate, and mean values of relative *Egr1* mRNA concentrations from three independent samples for buffer control, three independent samples for α -klotho^{ecto} treatment, and four independent samples for α -klotho^{ecto/ Δ RBA} treatment are shown in Extended Data Figs 1d and 7b, respectively. Protein elution profiles from size-exclusion columns shown in Figs 4a, 5a, b, f and Extended Data Fig. 7c are each representative of three independent experiments.

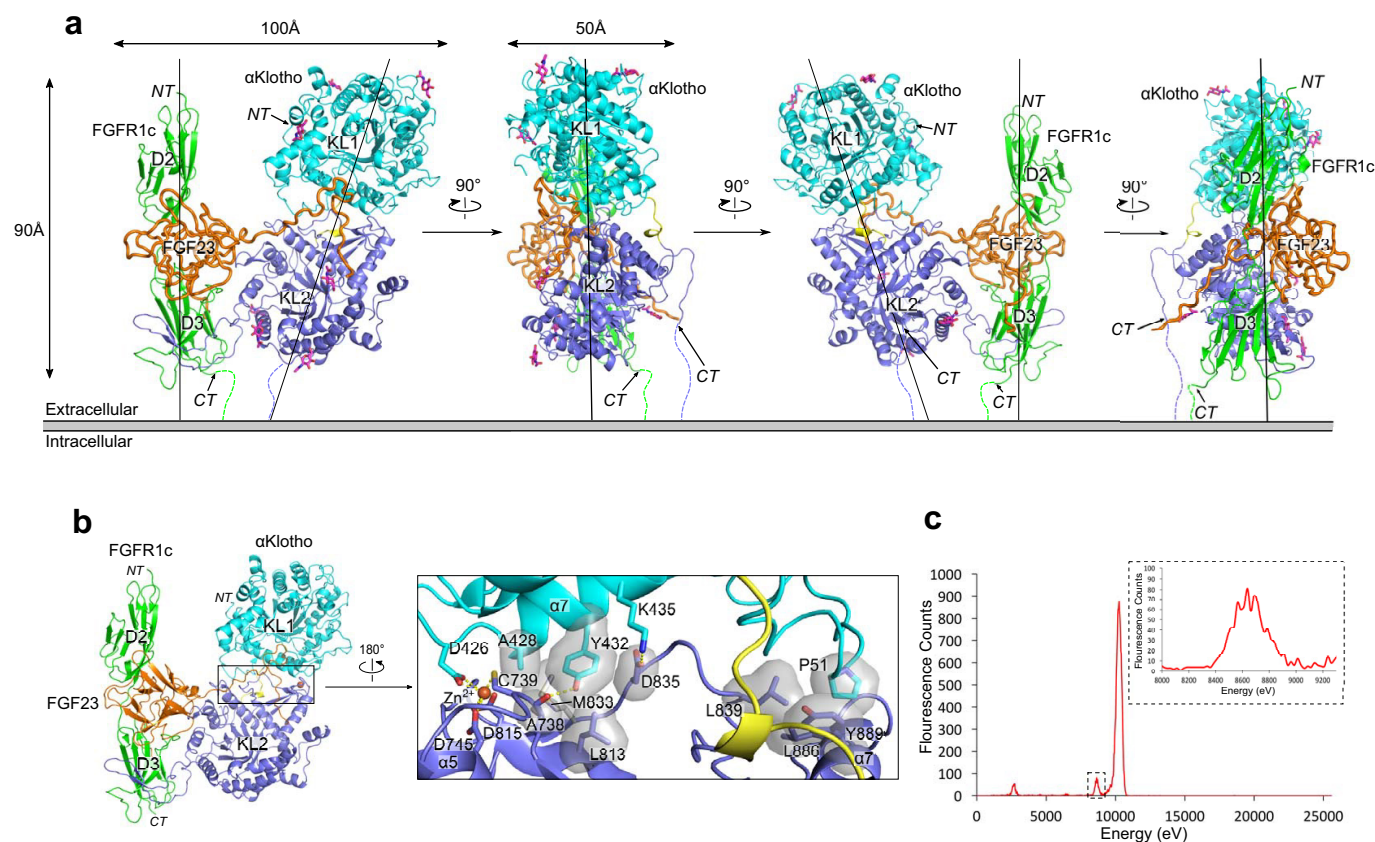
Data availability. Atomic coordinates and structure factors for the crystal structure of the FGF23–FGFR1c^{ecto}– α -klotho^{ecto} ternary complex are accessible at the RCSB Protein Data Bank (PDB) under accession code 5W21. Requests for *in vivo* datasets should be directed to O.W.M. Requests for all other reagents and datasets, including recombinant proteins, engineered cell lines and cell-based data, should be made to M.M.

37. White, K. E. *et al.* Autosomal dominant hypophosphataemic rickets is associated with mutations in *FGF23*. *Nat. Genet.* **26**, 345–348 (2000).
38. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
39. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
40. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
41. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
42. Goetz, R. *et al.* Molecular insights into the klotho-dependent, endocrine mode of action of fibroblast growth factor 19 subfamily members. *Mol. Cell. Biol.* **27**, 3417–3428 (2007).
43. Plotnikov, A. N., Schlessinger, J., Hubbard, S. R. & Mohammadi, M. Structural basis for FGF receptor dimerization and activation. *Cell* **98**, 641–650 (1999).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
45. Suzuki, M. *et al.* β Klotho is required for fibroblast growth factor (FGF) 21 signaling through FGF receptor (FGFR) 1c and FGFR3c. *Mol. Endocrinol.* **22**, 1006–1014 (2008).
46. Ornitz, D. M. *et al.* FGF binding and FGF receptor activation by synthetic heparan-derived di- and trisaccharides. *Science* **268**, 432–436 (1995).
47. Liu, Y. *et al.* Regulation of receptor binding specificity of FGF9 by an autoinhibitory homodimerization. *Structure* **25**, 1325–1336 (2017).
48. Belov, A. A. & Mohammadi, M. Molecular mechanisms of fibroblast growth factor signaling in physiology and pathology. *Cold Spring Harb. Perspect. Biol.* **5**, a015958 (2013).
49. Beenken, A., Eliseenkova, A. V., Ibrahimi, O. A., Olsen, S. K. & Mohammadi, M. Plasticity in interactions of fibroblast growth factor 1 (FGF1) N terminus with FGF receptors underlies promiscuity of FGF1. *J. Biol. Chem.* **287**, 3067–3078 (2012).
50. Olsen, S. K. *et al.* Structural basis by which alternative splicing modulates the organizer activity of FGF8 in the brain. *Genes Dev* **20**, 185–198 (2006).
51. Robinson, C. J., Harmer, N. J., Goodger, S. J., Blundell, T. L. & Gallagher, J. T. Cooperative dimerization of fibroblast growth factor 1 (FGF1) upon a single heparin saccharide may drive the formation of 2:2:1 FGF1.FGFR2c.heparin ternary complexes. *J. Biol. Chem.* **280**, 42274–42282 (2005).
52. Goodger, S. J. *et al.* Evidence that heparin saccharides promote FGF2 mitogenesis through two distinct mechanisms. *J. Biol. Chem.* **283**, 13001–13008 (2008).
53. Dunshee, D. R. *et al.* Fibroblast activation protein cleaves and inactivates fibroblast growth factor 21. *J. Biol. Chem.* **291**, 5986–5996 (2016).



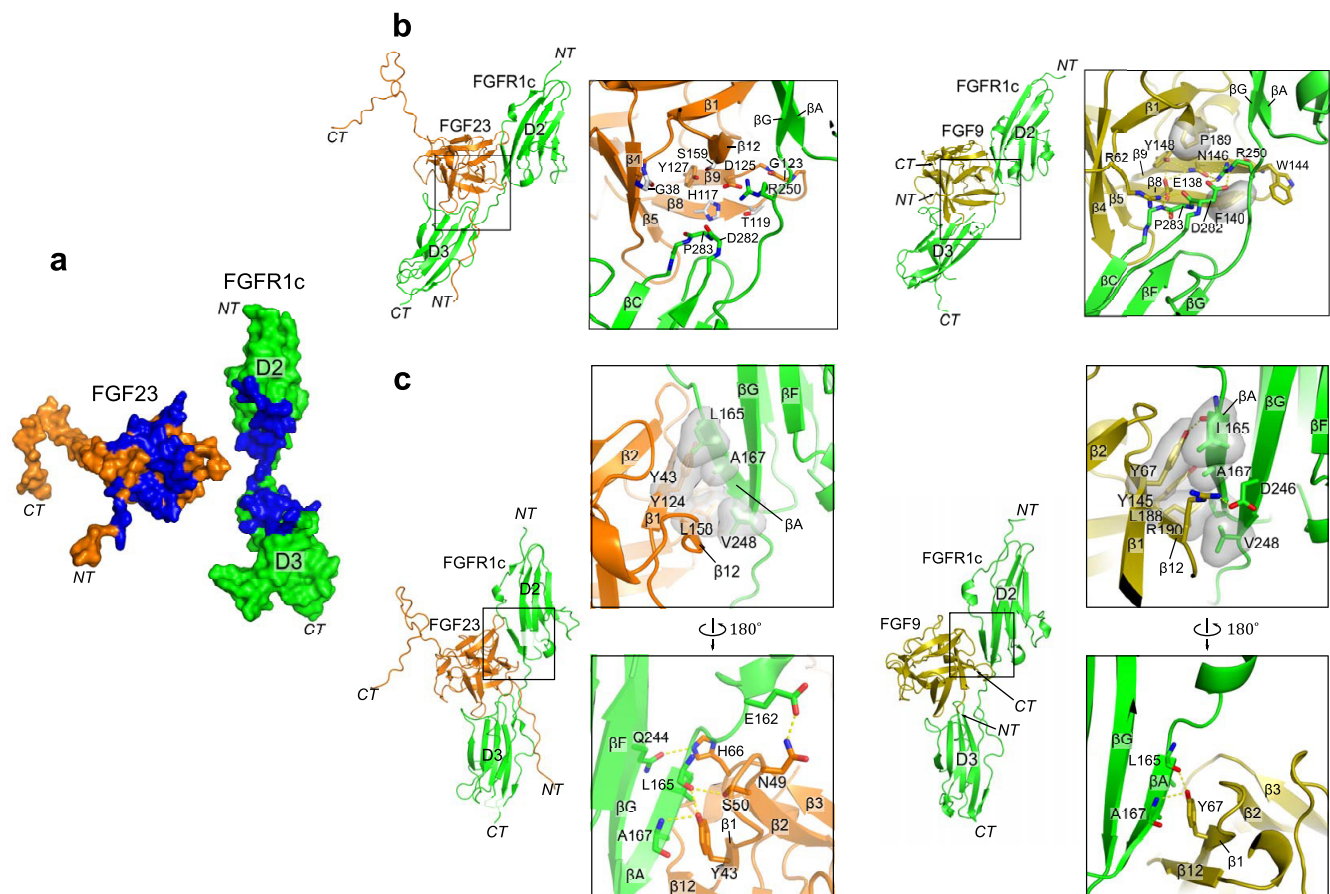
Extended Data Figure 1 | α -Klotho^{ecto} functions as a co-receptor for FGF23. **a**, Domain organization of membrane-bound α -klotho (α -klothoTM) and its soluble isoform α -klotho^{ecto} generated by an ectodomain shedding in the kidney¹⁶. KL1 and KL2 are tandem domains with homology to family 1 glycosidases⁸. **b**, Representative immunoblots of phosphorylated ERK (top) and total ERK (bottom; sample loading control) in total HEK293 cell lysates ($n = 3$ independent experiments). Top, lysates from untransfected HEK293 cells that were pre-treated with a fixed α -klotho^{ecto} concentration (10 nM) and then stimulated with increasing FGF23 concentrations, and lysates from HEK293- α -klothoTM cells treated with increasing concentrations of FGF23 alone.

Bottom, lysates from HEK293- α -klothoTM cells that were pre-treated with increasing α -klotho^{ecto} concentrations and then stimulated with a fixed FGF23 concentration. **c**, Plasma phosphate, fractional excretion of phosphate, and phosphate excretion rate in wild-type mice before and after a single injection of α -klotho^{ecto} (0.1 mg kg⁻¹ body weight) or isotonic saline alone (buffer). Circles denote mean values; error bars denote s.d. $n = 10$ mice per group. $*P < 0.05$, paired Student's t test. **d**, Relative *Egr1* mRNA levels in the kidney of wild-type mice after a single injection with α -klotho^{ecto} (0.1 mg kg⁻¹ body weight) or isotonic saline alone (buffer). Data are mean and s.d. $n = 3$ mice per group. The same batch of α -klotho^{ecto} protein was used in the experiments shown in **b–d**.



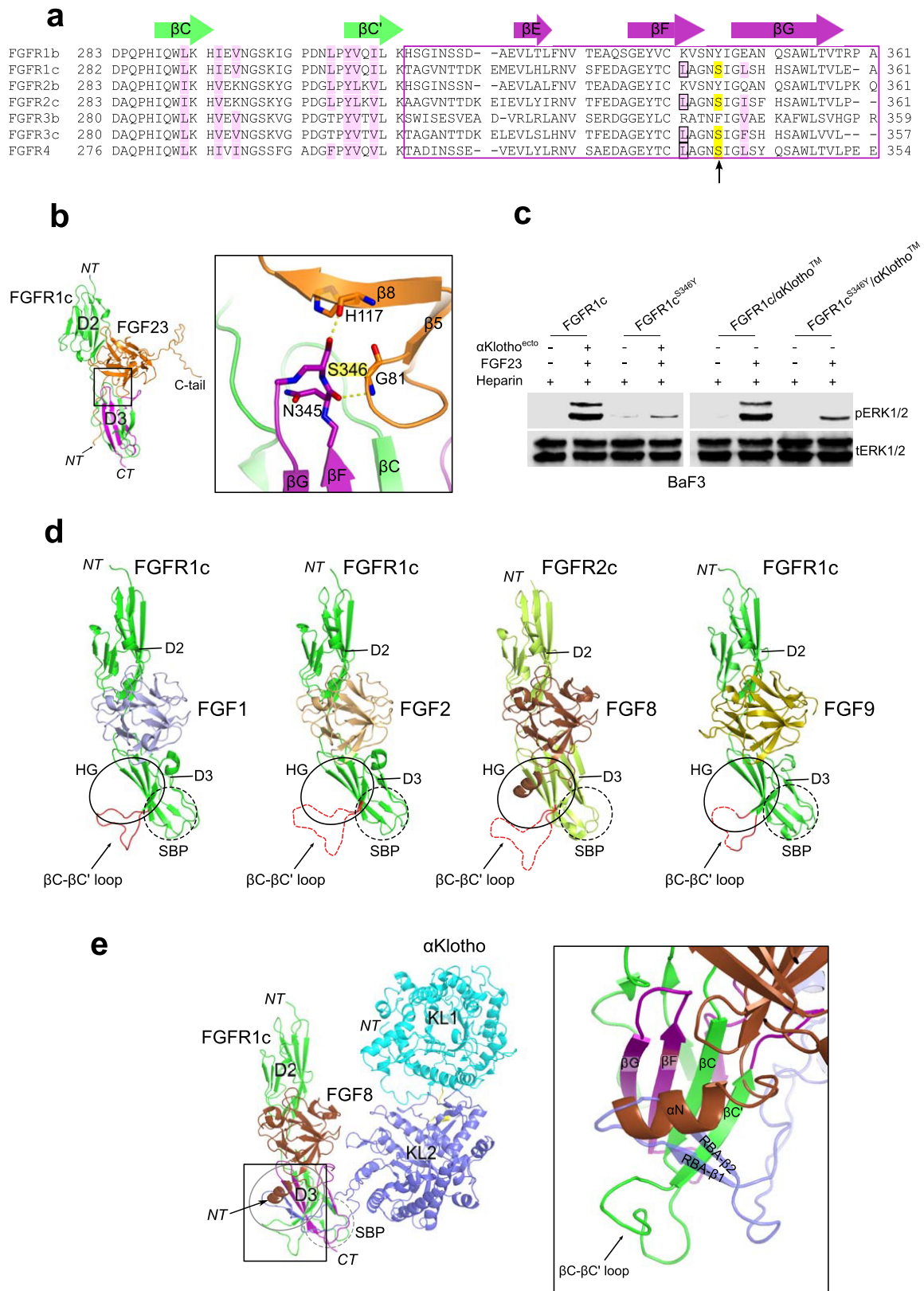
Extended Data Figure 2 | Topology of ternary complex is consistent with its orientation on the cell surface. **a**, Cartoon representation of 1:1:1 FGF23-FGFR1c^{ecto}-α-klotho^{ecto} complex in four different orientations related by 90° rotation. α-Klotho domains are coloured cyan (KL1) and blue (KL2); the KL1-KL2 linker is in yellow. FGFR1c and FGF23 are in green and orange, respectively. The ternary complex resembles an oblique rectangular prism with an average dimension of 100 Å × 90 Å × 50 Å. The long axes of α-klotho^{ecto} and FGF23-FGFR1c complex in the ternary complex are each about 90 Å long, and parallel to one another such that the C termini of FGFR1c^{ecto} and α-klotho^{ecto} end up on the same side of the ternary complex, ready to insert into the cell membrane (grey bar).

First, the *N*-acetyl glucosamine moiety (purple sticks) at six of the seven consensus *N*-linked α-klotho glycosylation sites could be built owing to sufficient electron density. Asn694 is the only glycosylation site that falls in the vicinity of a binding interface, namely α-klotho^{ecto}-FGF23. **b**, Close-up view of the KL1-KL2 interdomain interface. Zinc (orange sphere)-mediated contacts facilitate overall α-klotho^{ecto} conformation. Dashed yellow lines denote hydrogen bonds; grey surfaces denote hydrophobic contacts. **c**, Emission energy spectrum obtained from excitation/emission scan of the FGF23-FGFR1c^{ecto}-α-klotho^{ecto} crystal. Inset shows an expanded view of zinc fluorescence at 8,637 eV of emission energy.



Extended Data Figure 3 | Structural basis for the weak FGFR-binding affinity of FGF23. **a**, Open-book view of FGF23–FGFR1c^{ecto} complex interface. FGF23 (orange) and FGFR1c^{ecto} (green) are pulled apart and rotated by 90° around the vertical axis to expose the binding interface (blue). **b**, Ligand–receptor D3 and ligand–receptor D2–D3 linker interfaces of endocrine FGF23–FGFR1c and paracrine FGF9–FGFR1c⁴⁷ structures. Grey transparent surfaces denote hydrophobic interactions; dashed yellow lines denote hydrogen bonds. Because FGF9 Arg62 is replaced with glycine in FGF23 (Gly38) and FGF9 Glu138 is replaced with histidine in FGF23 (His117), neither the side chain of Asp125 in FGF23 (Asn146 in FGF9), nor the side chain of invariant Arg250 in the FGFR1c D2–D3 linker can be tethered through intramolecular hydrogen bonds. Thus, these side chains possess greater freedom of motion in the FGF23–FGFR1c complex, and as a result, hydrogen bonding between FGF23 and FGFR1c D2–D3 linker entails greater entropic cost, which

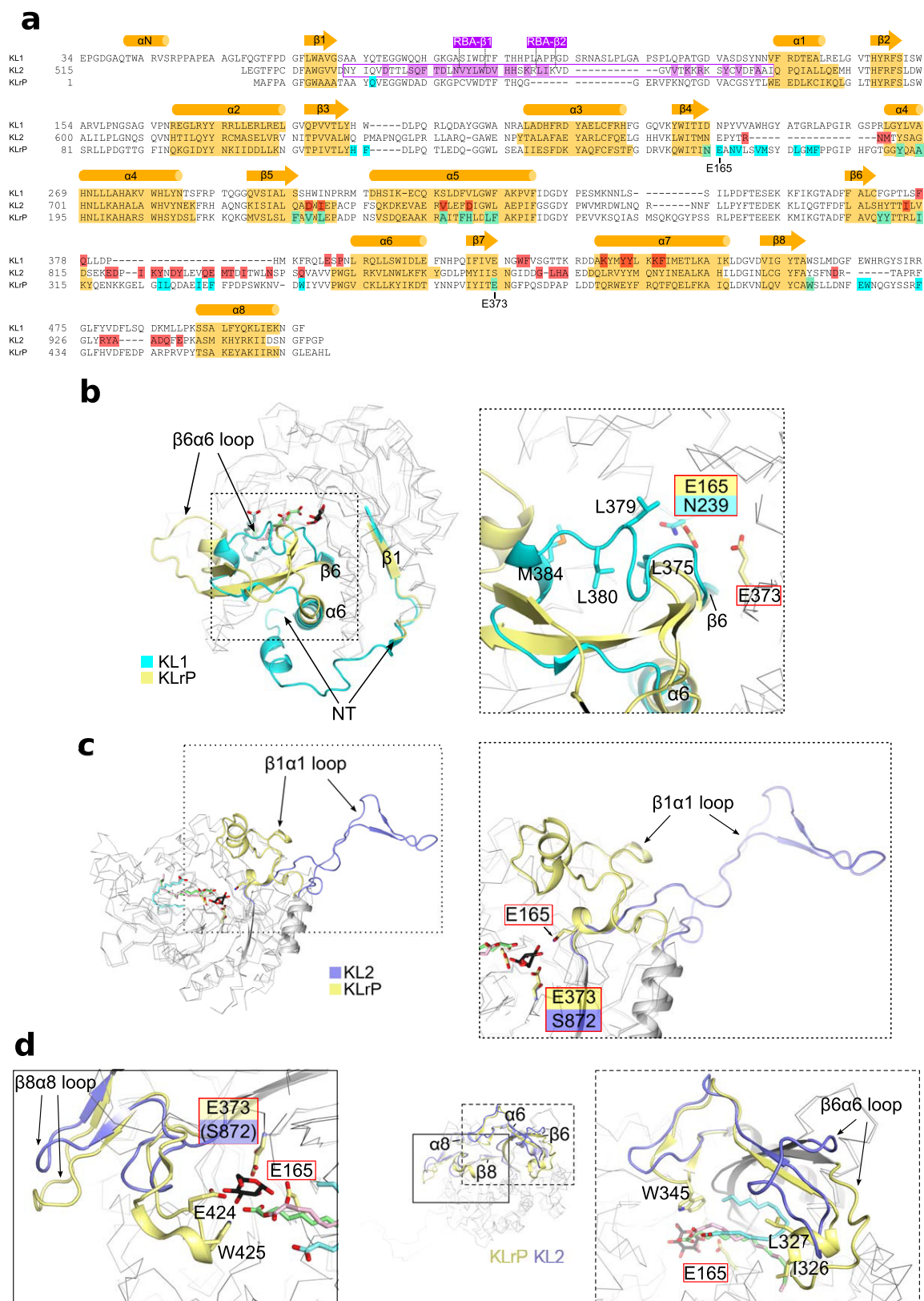
generates less binding affinity. Substitution of Phe140 and Pro189 in FGF9 with hydrophilic Thr119 and Ser159 in FGF23 further diminishes the ability of FGF23 to gain binding affinity from hydrogen bonding with FGFR1c D2–D3 linker. A lack of contacts between FGF23 N terminus and FGFR1c D3 cleft, which forms between alternatively spliced $\beta C'$ – βE and $\beta B'$ – βC loops⁴⁸, probably further exacerbates the weak FGFR-binding affinity of FGF23. **c**, Ligand–receptor D2 interface in endocrine FGF23–FGFR1c and paracrine FGF9–FGFR1c⁴⁷ structures. Grey transparent surfaces denote hydrophobic interactions; dashed yellow lines denote hydrogen bonds. Many contacts at this interface are conserved between paracrine FGF molecules and FGF23, and hence FGF23 gains much of its FGFR-binding affinity through these contacts. Three hydrogen bonds involving Asn49, Ser50 and His66 of FGF23 are unique to the FGF23–FGFR1c complex.



Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Structural basis for FGFR isoform specificity of α -klotho and FGF23. **a**, Structure-based sequence alignment of a segment of FGFR D3. The alternatively spliced regions of all seven FGFRs are boxed with a purple rectangle. β -strand locations above the alignment are coloured green (constant region) and purple (alternatively spliced region). A leucine (boxed) of hydrophobic groove residues (light purple) in the alternatively spliced region is conserved only among 'c' isoforms of FGFR1–FGFR3 and FGFR4, which explains α -klotho binding selectivity for these receptors. **b**, Interface between FGF23 and the β F– β G loop of FGFR1c D3 in the FGF23–FGFR1c structure of the ternary complex. Backbone atoms of His117 and Gly81 in FGF23 make specific hydrogen bonds with the Ser346 side-chain and Asn345 backbone atoms of the β F– β G loop. The serine residue corresponding to Ser346 in FGFR1c (yellow) is conserved only among 'c' isoforms of FGFR1–FGFR3 and FGFR4 (see **a**). **c**, Representative immunoblots of phosphorylated

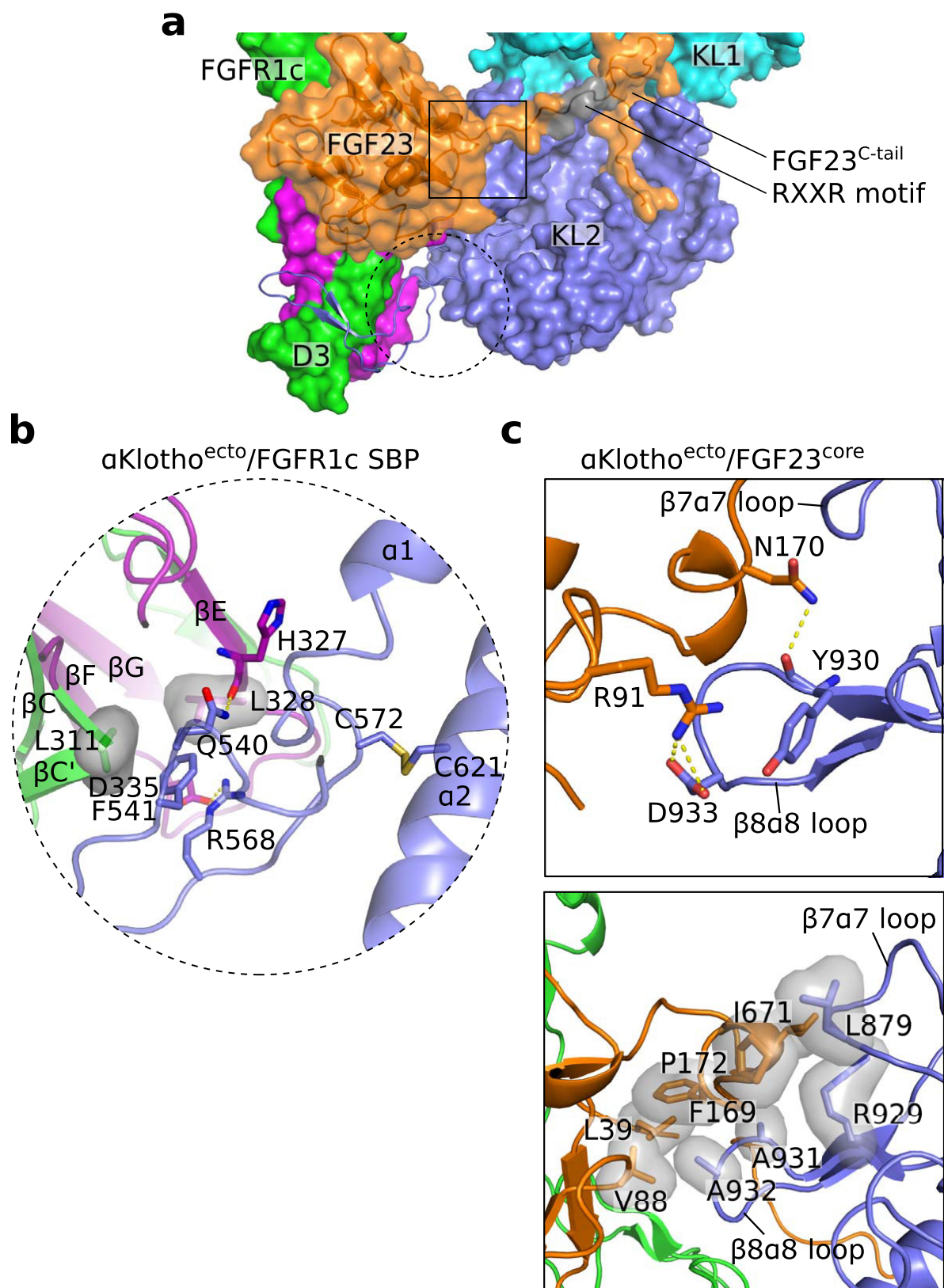
ERK (top) and total ERK (bottom; sample loading control) in total BaF3 cell lysates ($n = 3$ independent experiments). **d**, Cartoon representations of four paracrine FGF–FGFR complex structures^{4,47,49,50}. Solid black oval denotes the hydrophobic D3 groove. Dashed black circle denotes the second binding pocket (SBP) for α -klotho in D3. Although the hydrophobic groove is engaged by FGF8 (see also **e**), the SBP is not used in any of the current paracrine FGF–FGFR structures. In most paracrine FGF–FGFR structures, the β C– β C' loop is disordered (dashed red lines) because it does not participate in FGF binding. Evidently, SBP and β C– β C' loop in D3 have evolved to mediate α -klotho binding to FGFR. **e**, α -Klotho and FGF8b both bind to the hydrophobic groove in FGFR1c D3. FGF8b (brown) from the FGF8b–FGFR2c structure⁵⁰ was superimposed onto FGF23 in the FGF23–FGFR1c^{ecto}– α -klotho^{ecto} complex. The α N helix of FGF8b occupies the same binding pocket in FGFR1c D3 as the distal tip of the α -klotho RBA.



Extended Data Figure 5 | See next page for caption.

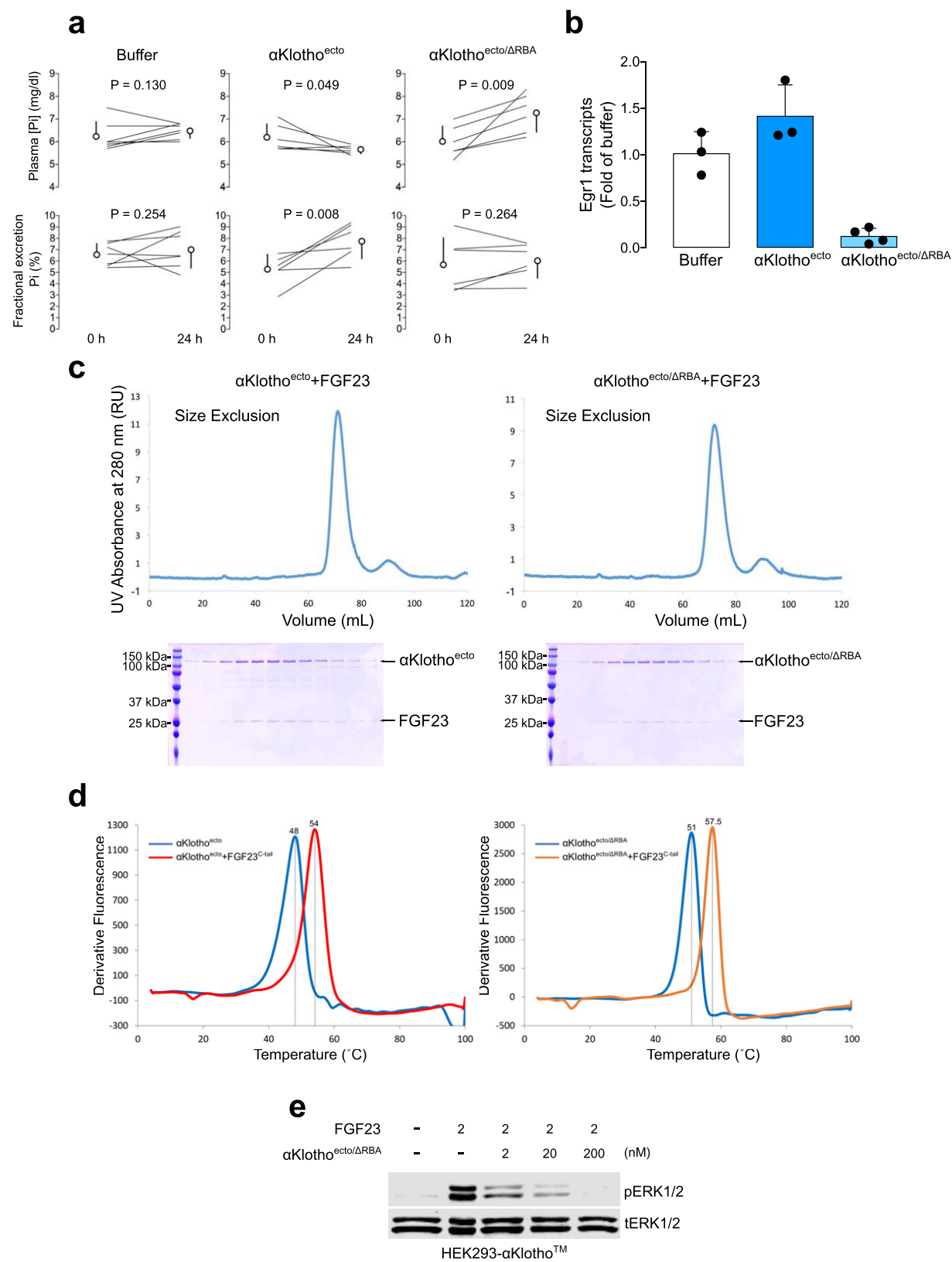
Extended Data Figure 5 | α -Klotho is the first non-enzymatic scaffold among TIM barrel proteins. **a**, Structure-based sequence alignment of TIM barrels of α -klotho KL1 and KL2 domains and KLRP. Most glycoside hydrolases (GH), a functionally diverse group of enzymes that cleave glycosidic bonds of complex carbohydrates on glycoproteins⁸, adopt a TIM barrel fold. Locations and lengths of TIM barrel β -strands and α -helices are indicated above the alignment. Among GH family 1 members of the klotho subfamily, only KLRP has a verified glycosylceramidase activity²⁶, and Glu165 and Glu373 are its catalytically essential glutamic acids. KLRP residues coloured cyan participate in substrate recognition/hydrolysis. α -Klotho residues coloured red bind FGF23, and α -klotho residues of the KL2 β 1 α 1 loop (purple box) coloured purple interact with the FGFR1c D3 domain. **b**, Superimposition of KL1 C α trace (grey/cyan) onto that of KLRP (grey/yellow). Superimposition root mean square deviation (r.m.s.d.) value is 1.08 Å. Structurally most divergent regions between KL1 and KLRP are in cartoon representation. Glucose moiety and aliphatic chains of glucosylceramide (KLRP substrate) are in sticks with carbon in black (glucose) or green/cyan/pink (aliphatic chains). Catalytically essential Glu165 in KLRP is replaced by an asparagine in KL1. Hydrophobic residues

from KL1 β 6 α 6 loop occupy the pocket that accommodates the aliphatic chains of glucosylceramide in KLRP. The KL1 N terminus supports KL1–KL2 cleft formation (Extended Data Fig. 2b) and KL1 β 6 α 6 loop conformation contributes to a key portion of the binding pocket in this cleft for the FGF23 C-terminal tail (Fig. 3c). **c, d**, Superimposition of KL2 C α trace (grey/blue) onto that of KLRP (grey/yellow). Superimposition r.m.s.d. value is 1.37 Å. Structurally divergent β 1 α 1 (**c**), β 6 α 6 and β 8 α 8 (**d**) loops of KL2 and KLRP are rendered in cartoon. β 1 α 1 loop in KL2 is disengaged from the central TIM barrel and stretches away from it by as much as 35 Å. Catalytically essential Glu373 in KLRP is replaced by a serine in KL2. KLRP residues from β 6 α 6 and β 8 α 8 loops bind glucosylceramide (KLRP substrate); for example, Trp345 in the β 6 α 6 loop and Glu424 and Trp425 in the β 8 α 8 loop. Sequence divergence (**a**) and altered loop conformations are incompatible with glucosylceramide coordination by KL2. β 1 α 1, β 6 α 6 and β 8 α 8 loops lie at the rim of the catalytic mouth in the TIM barrel (see Fig. 2b). Divergent conformations of these three loops in KL2 result in notable widening of the central barrel cavity in KL2, which merges with the KL1–KL2 cleft to form an expansive basin that accommodates the distal portion of the FGF23 C-terminal tail.



Extended Data Figure 6 | α -Klotho interaction with rigid core of FGF23 and a second binding pocket next to the hydrophobic groove in FGFR1c D3. **a**, A partial view of the ternary complex. α -Klotho^{ecto} (cyan/blue solid surface, RBA of KL2 in blue cartoon), FGF23 (orange transparent surface and cartoon), FGFR1c (constant region: solid green surface; alternatively spliced region: solid purple surface). Dashed black circle denotes the perimeter of the interface between proximal end of α -klotho RBA and a second binding pocket (SBP) in FGFR1c D3 next to the hydrophobic groove. Solid black box denotes the perimeter of

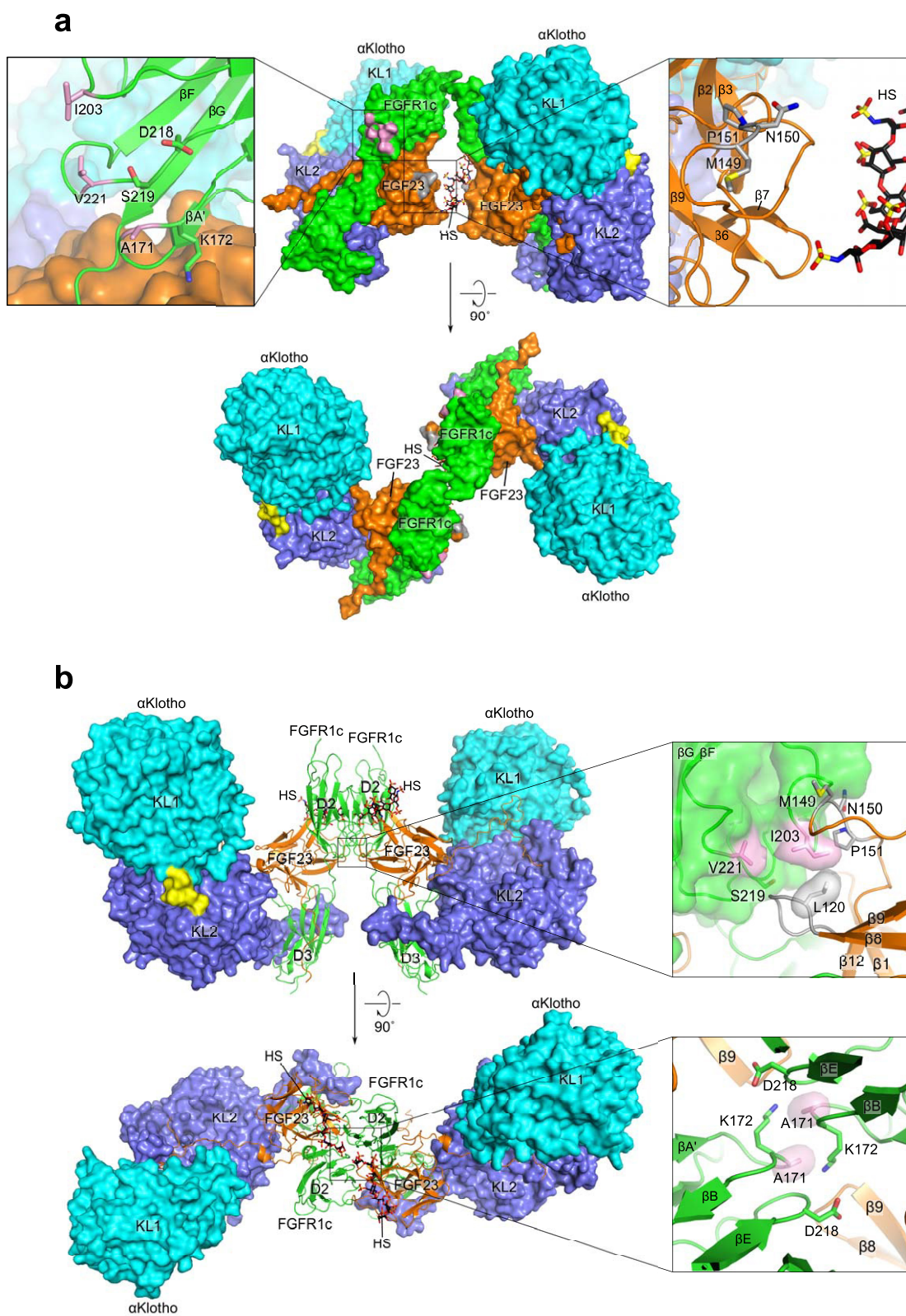
α -klotho–FGF23^{core} interface. **b**, Close-up view of the interface between proximal end of RBA and SBP in D3. The disulfide bridge between Cys572 (N-terminal end of RBA) and Cys621 (α 2 helix) at the base of the RBA probably imparts some degree of conformational rigidity to the proximal RBA portion, whereas the conformation of the distal RBA tip is dictated by contacts with FGFR1c D3. **c**, Close-up view of the α -klotho–FGF23^{core} interface detailing hydrogen bonding (top) and hydrophobic contacts (bottom). Grey transparent surfaces denote hydrophobic interactions; dashed yellow lines denote hydrogen bonding contacts.



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Deletion of RBA of α -klotho^{ecto} generates an FGF23 ligand trap. **a**, Plasma phosphate and fractional excretion of phosphate in wild-type mice before and after a single injection of α -klotho^{ecto} (0.1 mg kg⁻¹ body weight), mutant α -klotho^{ecto/ Δ RBA} (0.1 mg kg⁻¹ body weight), or isotonic saline alone (buffer). Circles denote mean values; error bars denote s.d. $n = 6$ mice per group. Significance values were determined by a paired Student's t test. **b**, Relative *Egr1* mRNA levels in the kidney of wild-type mice injected once with α -klotho^{ecto} (0.1 mg kg⁻¹ body weight; $n = 3$), mutant α -klotho^{ecto/ Δ RBA} (0.1 mg kg⁻¹ body weight; $n = 4$), or isotonic saline alone (buffer; $n = 3$). Data are mean and s.d. **c**, Representative elution profiles of FGF23– α -klotho^{ecto} and FGF23– α -klotho^{ecto/ Δ RBA} mixtures from a size-exclusion column and representative Coomassie blue-stained SDS-polyacrylamide gels of eluted protein peak fractions. **d**, Thermal shift assay of α -klotho^{ecto} and the α -klotho^{ecto/ Δ RBA} mutant in the presence and absence of FGF23

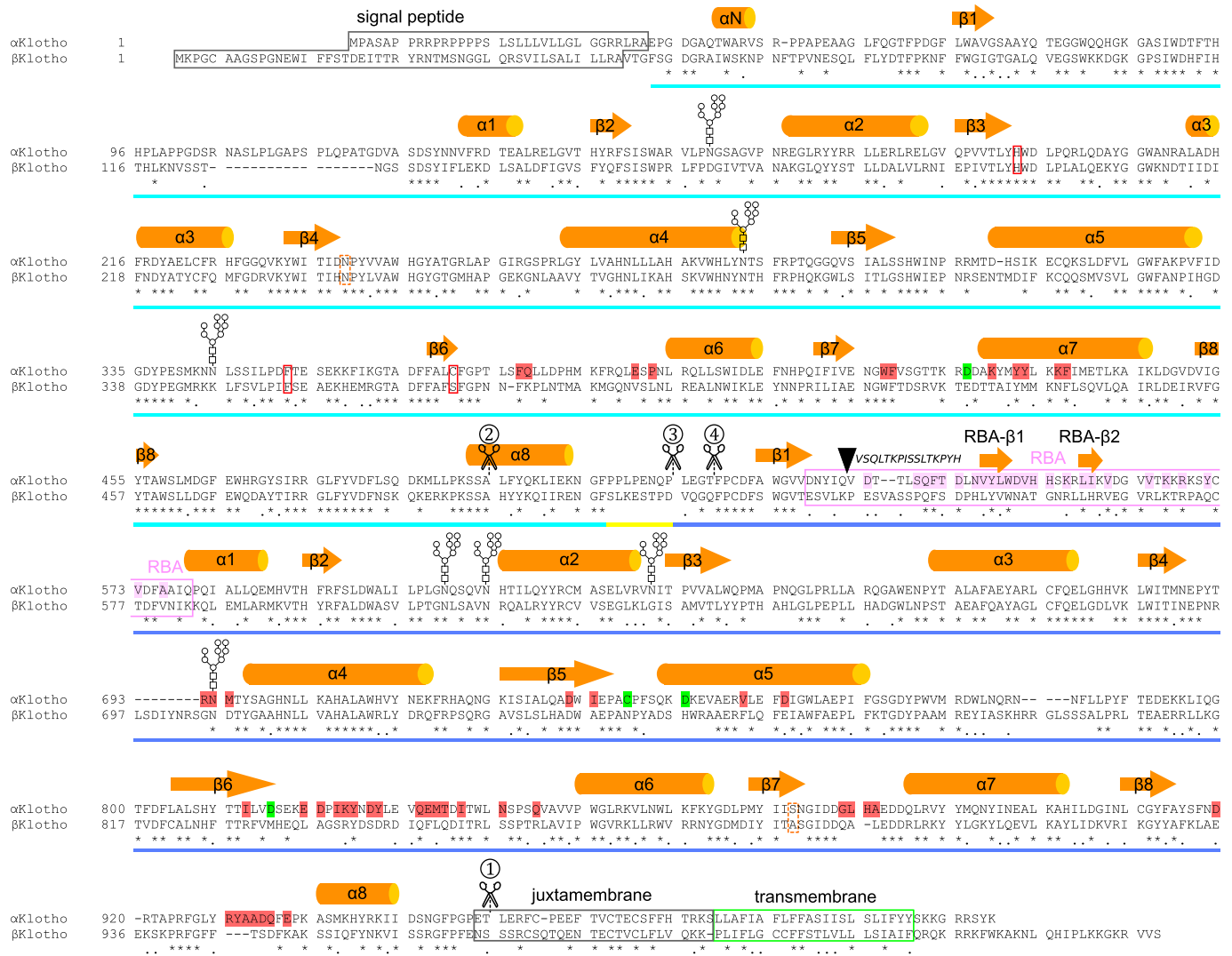
C-terminal tail peptide (FGF23^{C-tail}) ($n = 3$ independent experiments). Increased melting temperatures in the presence of the FGF23^{C-tail} indicate interaction of both α -klotho^{ecto} proteins with the peptide. Higher melting temperature of α -klotho^{ecto/ Δ RBA} mutant relative to wild-type α -klotho^{ecto} indicates greater stability of the mutant protein. **e**, Representative immunoblots of phosphorylated ERK (top) and total ERK (bottom; sample loading control) in total lysates from HEK293– α -klothoTM cells co-stimulated with a fixed FGF23 concentration and increasing α -klotho^{ecto/ Δ RBA} concentrations ($n = 3$ independent experiments). The α -klotho^{ecto/ Δ RBA} mutant inhibits FGF23-induced ERK phosphorylation owing to sequestering FGF23 into inactive FGF23– α -klotho^{ecto/ Δ RBA} binary complexes. This also explains why α -klotho^{ecto/ Δ RBA} injection into mice causes an increase in plasma phosphate (**a**) concomitant with renal *Egr1* gene repression (**b**).



Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | FGF23–FGFR1^{ecto}– α -klotho^{ecto}–HS quaternary dimer models. **a**, A 2:2:2:1 FGF23–FGFR1^{ecto}– α -klotho^{ecto}–HS quaternary dimer in two orientations related by a 90° rotation around the horizontal axis. The dimer was constructed by superimposing FGF23 from two copies of 1:1:1 FGF23–FGFR1^{ecto}– α -klotho^{ecto} complex onto the two FGF1 molecules in the 2:2:1 FGF1–FGFR2c–HS dimer^{30,31,51,52}. The dimer is held together solely by HS, which bridges two FGF23 molecules in *trans*. Boxed pink surface denotes the location of Ala171, Ile203 and Val221 of FGFR1c, the mutation of which impairs the ability of HS to induce 2:2:2:2 quaternary dimer formation (Fig. 5f). Boxed grey region denotes the location of Met149, Asn150 and Pro151 of FGF23, the mutation of which diminishes HS-induced quaternary dimerization (Fig. 5e, f). None of these residues has any role in 2:2:2:1 quaternary dimer

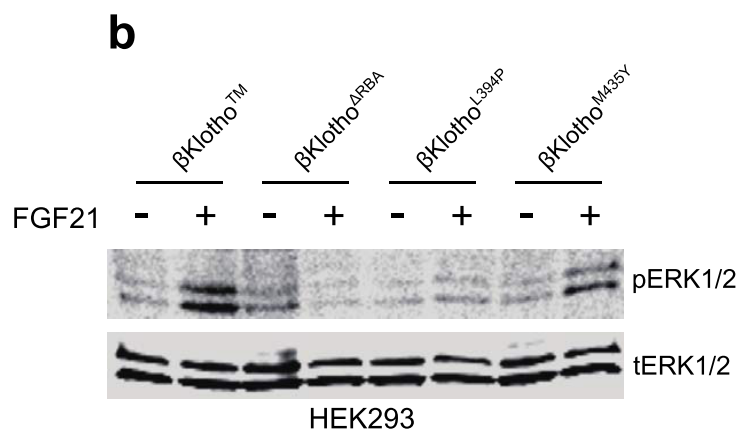
formation, and hence, contrary to experimental evidence (Fig. 5), mutation of these residues should not affect HS-induced FGF23–FGFR1^{ecto}– α -klotho^{ecto} dimerization. **b**, A 2:2:2:2 FGF23–FGFR1^{ecto}– α -klotho^{ecto}–HS quaternary dimer in two orientations related by a 90° rotation around the horizontal axis. See also Fig. 5g. The dimer was constructed by superimposing FGF23 from two copies of 1:1:1 FGF23–FGFR1^{ecto}– α -klotho^{ecto} complex onto the two FGF2 molecules in the 2:2:2 FGF2–FGFR1c–HS dimer⁴. Insets show close-up views of the secondary FGF–FGFR (top) and direct FGFR–FGFR (bottom) interfaces. Grey/pink transparent surfaces denote hydrophobic interactions. Mutation of Ala171, Ile203 and Val221 (pink) impairs the ability of HS to dimerize the FGF23–FGFR1^{ecto}– α -klotho^{ecto} ternary complex (Fig. 5f).



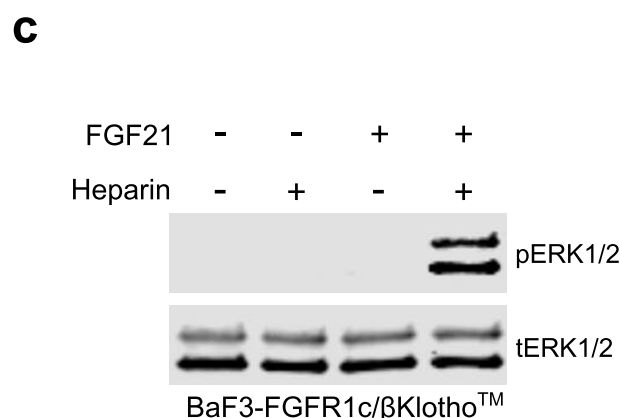
Extended Data Figure 9 | The FGF19 and FGF21 co-receptor β-klotho is a non-enzymatic scaffold protein analogous to α-klotho. Structure-based sequence alignment of α-klotho and β-klotho. The locations of the eight alternating β-strands and α-helices of the TIM fold are indicated above the alignment. Cyan, blue and yellow bars below the alignment mark the domain boundaries of KL1, KL2 and the KL1–KL2 linker. Asterisks denote sequence identity and dots denote sequence similarity. Scissor symbols mark the four proposed sites of α-klotho cleavage by ADAM proteases/secretases. Cleavage 1, which coincides with the end of the rigid core of KL2, results in shedding of the entire α-klotho ectodomain from the cell membrane. Although this cleavage product is a functional co-receptor, the α-klotho fragments generated by cleavages 2, 3 and 4 would be devoid of co-receptor activity. Black triangle denotes the site

where alternative splicing replaces the C-terminal KL2 sequence with a 15-residue-long unrelated sequence. Glycan chain symbols denote seven predicted N-linked glycosylation sites. Zn²⁺-chelating residues of α-klotho are green, FGFR1c-binding residues are light purple, and FGF23-binding residues are red. Light purple box denotes β1α1 loop sequence in KL2 termed RBA. β-Klotho RBA is about as long as α-klotho RBA, and key FGFR-binding residues are conserved between these two RBAs, which is consistent with the similar FGFR-binding specificity of α-klotho and β-klotho^{9,11,12}. But α-klotho residues in the binding pockets for the FGF23 C-terminal tail are not conserved in β-klotho, conforming to major sequence differences between the C-terminal tails of FGF23, FGF19 and FGF21 (Extended Data Fig. 10a).

Protein	Residue	Sequence
FGF23	25	YPN ASPLLGGSSWG GLIHLYTATA RN--SYHLQI HKNGHVDGAP HQTIYSALMI
FGF21	29	HPIDPS SPLLQFGGQV RQRYLYTDDA QQ-TEAHLEI REDGTVGGAA DQSPESLLQL
FGF19	23	RPLAFSDAGP HVHYGWGDPI RLRHLYTSGP HGLSSCFLRI RADGVVDCAR GQSAHSLLEI
	 * * * * * * . . * * . . . *
FGF23	76	RSEDAGFVVI TGVMSRRYLC MDFRGNIFGS HYFDPENCRF QHQTLENGYD VYHSPQYHFL
FGF21	84	KALKPGVIQI LGVKTSRFLC QRPDGALYGS LHFDPACSF RELLEDGYN VYQSEAHGLP
FGF19	83	KAVALRTVAI KGVHSVRYLC MGADGKMQL LQYSEEDCAF EEEIRPDGYN VYRSEKHLRP
	 * * * . . * * * * * * *
FGF23	136	VSLGRAKRAF LP-GMNPPPY SQFLSRNEI PLIHFNTP PIP RGHTRSAEDD SERDPLNLVK
FGF21	144	LHLPGNKSPH RD--PAPRGP ARFLPLPGLP PALPEPPGIL APQPPDVGSS DPLSMVG---
FGF19	143	VSLSSAKQRQ LYKNRGFLPL SHFLPMLPMV PEEPEDLRGH LESDMFSSPL ETDSMDPFGL
		. . * * * * *
FGF23	195	PRAMT PAPA SCSQELPSAE DNSPMASDPL GVVRRGRVNT HAGGTGPEGC RPF AKFI
FGF21	199	---PSQGRSP SYAS
FGF19	203	VTGLEAVRSP SEEK
		* *



Extended Data Figure 10 | β -Klotho-dependent FGFR activation by FGF19 and FGF21 is mechanistically similar to α -klotho-dependent FGFR activation by FGF23. **a**, Structure-based sequence alignment of endocrine FGF proteins. β -strands and the α C helix comprising the atypical β -trefoil core of FGF23 are indicated above the alignment. Asterisks and dots below the alignment denote sequence identity and similarity, respectively. Scissor symbols mark inactivating proteolytic cleavage sites in FGF23 and FGF21⁵³. RXXR cleavage motif in FGF23 is in green bold letters. FGFR1c-binding residues of FGF23 are coloured blue, α -klotho-binding residues are coloured red. Vertical blue arrow marks the C-terminal boundary of the FGF23 variant used to solve the FGF23-FGFR1c^{ecto}- α -klotho^{ecto} complex structure. Five residues at the distal C-terminal region of FGF19 or FGF21 (black and grey) mediate binding of FGF19 or FGF21 to β -klotho. These residues completely diverge from the α -klotho-binding residues in the FGF23 C-terminal tail.



α -Klotho-binding residues in the FGF23 core also are not conserved in FGF19 and FGF21. **b**, Representative immunoblots of phosphorylated ERK (top) and total ERK (bottom; sample loading control) in total lysates from HEK293 cells expressing wild-type or mutant β -klothoTM ($n = 3$ independent experiments). Similar to α -klotho Δ RB Δ , β -klotho Δ RB Δ failed to support FGF21-induced FGFR activation, and β -klotho (L394P) and β -klotho (M435Y) mutants also had greatly diminished ability to promote FGF21 signalling. Thus, β -klotho tethers FGFR1c and FGF21 to itself in a manner similar to that identified for α -klotho to enable FGF21 signalling. **c**, Representative immunoblots of phosphorylated ERK (top) and total ERK (bottom; sample loading control) in total lysates from BaF3 cells expressing FGFR1c and β -klothoTM ($n = 3$ independent experiments). Like α -klotho, β -klotho also requires heparin to support FGF21-mediated FGFR1c activation.

Extended Data Table 1 | X-ray data collection and structure refinement statistics

Protein	FGF23-FGFR1c ^{ecto} - α Klotho ^{ecto}
Data Collection	
X-ray wavelength (Å)	0.97918
Space group	C2
Unit Cell Dimensions	
a, b, c (Å)	283.31, 72.60, 95.33
α , β , γ (°)	90.00, 91.98, 90.00
Resolution (Å)	50-3.00 (3.18-3.0)
No. measured reflections	294862
No. unique reflections	39077
Data redundancy	7.5 (7.6)
Data completeness (%)	99.7 (98.8)
R _{meas} (%)	20.7 (138.0)
Signal (<I/>σI)	11.1 (1.7)
Refinement	
Resolution (Å)	48.81-3.00 (3.08-3.00)
No. unique reflections	38950 (2688)
No. reflections (R _{free})	1947 (133)
R _{work} /R _{free}	23.00 (44.46)/27.82 (51.89)
No. TLS groups	3 (one per polypeptide chain)
Number of atoms	
Protein	10602
Sugar (NAG)	98
Ion (Zn ²⁺)	1
Solvent	1
R.m.s. deviations	
Bond length (Å)	0.002
Bond angle (°)	0.483
Average B factors (Å²)	
Protein	114
Sugar (NAG)	180
Ion (Zn ²⁺)	116
Solvent	58
Ramachandran Plot	
Favored (%)	89.06
Allowed (%)	9.72
Outliers (%)	1.22
Rotamer outliers (%)	2.23
No. C β Deviations	0
All-Atom Clashscore	6.5
PDB ID	5W21

Values in parentheses are for the highest resolution shell.

Chromosomal instability drives metastasis through a cytosolic DNA response

Samuel F. Bakhom^{1,2*}, Bryan Ngo^{2*}, Ashley M. Laughney³, Julie-Ann Cavallo^{1,2}, Charles J. Murphy², Peter Ly⁴, Pragma Shah⁵, Roshan K. Sriram², Thomas B. K. Watkins⁶, Neil K. Taunk¹, Mercedes Duran^{1,2}, Chantal Pauli⁷, Christine Shaw⁸, Kalyani Chadalavada⁸, Vinagolu K. Rajasekhar⁹, Giulio Genovese¹⁰, Subramanian Venkatesan¹¹, Nicolai J. Birkbak^{6,11}, Nicholas McGranahan^{6,11}, Mark Lundquist², Quincey LaPlant¹, John H. Healey⁹, Olivier Elemento², Christine H. Chung¹², Nancy Y. Lee¹, Marcin Imielenski², Gouri Nanjangud⁸, Dana Pe'er¹³, Don W. Cleveland⁴, Simon N. Powell¹, Jan Lammerding⁵, Charles Swanton^{6,11} & Lewis C. Cantley²

Chromosomal instability is a hallmark of cancer that results from ongoing errors in chromosome segregation during mitosis. Although chromosomal instability is a major driver of tumour evolution, its role in metastasis has not been established. Here we show that chromosomal instability promotes metastasis by sustaining a tumour cell-autonomous response to cytosolic DNA. Errors in chromosome segregation create a preponderance of micronuclei whose rupture spills genomic DNA into the cytosol. This leads to the activation of the cGAS–STING (cyclic GMP–AMP synthase–stimulator of interferon genes) cytosolic DNA-sensing pathway and downstream noncanonical NF- κ B signalling. Genetic suppression of chromosomal instability markedly delays metastasis even in highly aneuploid tumour models, whereas continuous chromosome segregation errors promote cellular invasion and metastasis in a STING-dependent manner. By subverting lethal epithelial responses to cytosolic DNA, chromosomally unstable tumour cells co-opt chronic activation of innate immune pathways to spread to distant organs.

Chromosomal instability (CIN) correlates with tumour metastasis^{1,2}, but it remains unclear whether it is a mere bystander or a driver of metastatic progression. Chromosomally unstable cells show evidence of chromosome missegregation during anaphase^{3,4}, offering an attractive bottleneck in which to target CIN and probe its selective contribution to metastasis. Destabilization of microtubule attachments to chromosomes at the kinetochores, through overexpression of the non-motile microtubule-depolymerizing kinesin-13 family proteins KIF2B or KIF2C (also known as MCAK), directly suppresses CIN in otherwise chromosomally unstable cells^{5–7}. Cells overexpressing KIF2B or MCAK continue to propagate abnormal aneuploid karyotypes, albeit in a stable manner⁷. As such, this approach permits direct experimental interrogation of CIN, as defined by the rate of ongoing chromosome missegregation, independently of aneuploidy, which is defined as a state of abnormal chromosome numbers.

Increased CIN in human metastases

First, to determine whether CIN is associated with human metastases, we applied the weighted-genomic integrity index (wGII) as a proxy for CIN⁸ to 79 matched pairs of primary tumour and brain metastasis from a published cohort⁹. Metastases showed higher wGII than primary tumours (Fig. 1a and Extended Data Fig. 1a, b).

Next, karyotype analysis of primary breast tumours and metastases archived in the Mitelman Database of chromosomal translocations¹⁰ revealed a predilection for near-diploid ($2n$) karyotypes in primary tumours. Conversely, metastases were enriched for cells with near-triploid ($3n$) karyotypes and had twice as many structural or numerical

chromosomal aberrations per clone as primary tumours. The number of chromosomal aberrations was highest in tumour samples with karyotypes ranging between the diploid and tetraploid ($4n$) range (Fig. 1b, c and Extended Data Fig. 1c, d).

Finally, histological analysis of primary tumours from patients with locally advanced head and neck squamous cell carcinoma¹¹ revealed a significant association ($P < 0.05$) between anaphase chromosome missegregation and the incidence of lymph node metastasis (Fig. 1d and Extended Data Fig. 1e).

CIN is a driver of metastasis

To determine whether CIN is causally involved in metastasis, we used transplantable metastatic tumour models of human (MDA-MB-231) or mouse (4T1) triple-negative breast cancer and human lung adenocarcinoma (H2030), in which 47%, 55%, and 67% of anaphase cells, respectively, show evidence of chromosome missegregation. Overexpression of KIF2B or MCAK suppressed chromosome missegregation, whereas overexpression of a dominant-negative MCAK mutant¹² (dnMCAK) led to a modest increase in chromosome missegregation in MDA-MB-231 cells. Overexpression of KIF2B or MCAK did not alter cellular proliferation or the number of centrosomes per cell (Fig. 2a, b and Extended Data Fig. 1f–i). As a control, we overexpressed KIF2A, a third member of the kinesin-13 family that lacks kinetochore and centromere localization domains¹³; although KIF2A showed microtubule-depolymerizing activity on interphase microtubules, it had no observable effect on CIN (Fig. 2b and Extended Data Fig. 1i–k). We ruled out a direct role for kinesin-13-mediated microtubule depolymerization in

¹Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ²Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, New York 10065, USA. ³Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ⁴Ludwig Institute for Cancer Research, University of California San Diego, La Jolla, California 92093, USA. ⁵Nancy E. and Peter C. Meinig School of Biomedical Engineering & Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, New York 14850, USA. ⁶The Francis Crick Institute, London NW1 1AT, UK. ⁷Institute for Pathology and Molecular Pathology, University Hospital Zurich, Zurich 8091, Switzerland. ⁸Molecular Cytogenetics Core, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ⁹Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ¹⁰The Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ¹¹UCL Cancer Institute, London WC1E 6BT, UK. ¹²Moffitt Cancer Center, Tampa, Florida 33612, USA.

¹³Computational Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA.

*These authors contributed equally to this work.

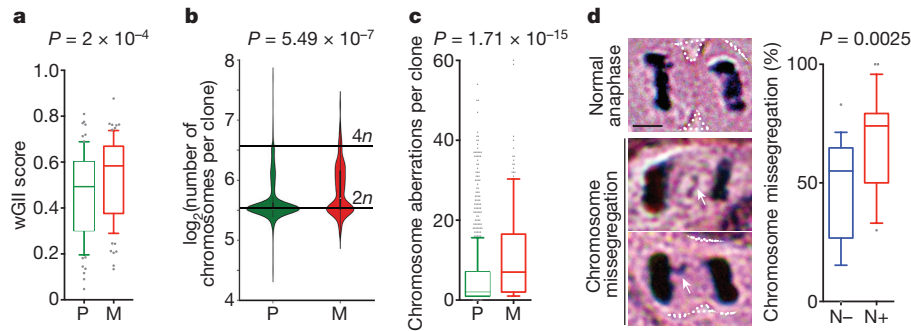


Figure 1 | Human metastases enrich for CIN. **a**, wGII of matched primary tumours (P) and brain metastases (M), $n = 79$ patients. **b**, **c**, Karyotype probability density (**b**) and chromosomal aberrations (**c**) in 983 primary tumour and 186 metastatic breast cancer clones. **d**, Left, images of head and neck squamous cell carcinoma cells undergoing anaphase. Arrows point to chromosome missegregation; scale bar, 5 μ m.

Right, chromosome missegregation in tumours from patients with (N+, $n = 22$ patients) or without (N-, $n = 18$ patients) clinically detectable lymph node metastases. Boxes represent median \pm interquartile range, confidence intervals (whiskers) denote 10th–90th percentiles (**a**, **c**, **d**), significance tested using two-sided Wilcoxon matched-pairs signed rank test (**a**) and two-sided Mann–Whitney test (**b**–**d**).

activating small GTPases¹⁴ by performing RhoA and Rac1 pull-down assays, which revealed low basal levels of activity and no correlation with overexpression of kinesin-13 family proteins (Extended Data Fig. 2a, b). Hereafter, we refer to cells expressing MCAK or KIF2B as CIN-low and to control cells or those expressing KIF2A or dnMCAK as CIN-high.

Karyotyping of the parental MDA-MB-231 cells revealed a widely aneuploid (approximately $3n$) chromosome content with widespread karyotypic heterogeneity (Extended Data Fig. 2c). Suppression of CIN reduced both numerical and structural karyotypic heterogeneity in single-cell-derived clones, as supported by the presence of fewer chromosomes exhibiting non-clonal structural abnormalities and decreased numerical chromosome heterogeneity in CIN-low cells (Extended Data Fig. 2c–h). Notably, CIN-low cells maintained highly aneuploid karyotypes, but faithfully propagated them in a stable manner. Thus, by comparing chromosomally stable aneuploid cells to their chromosomally unstable aneuploid counterparts, we can experimentally examine the role of CIN, independent of aneuploidy, in metastasis.

We injected MDA-MB-231 cells into the left cardiac ventricles of athymic mice to enable systemic dissemination while tracking metastatic colonization using a bioluminescence reporter. Differences in chromosome missegregation rates had a marked effect on colonization: mice harbouring CIN-high cells rapidly succumbed to metastatic disease, with a median survival of 70 days, whereas mice injected with CIN-low cells had a lower metastatic burden and a median survival of 207 days. Many metastases from CIN-low cells waxed and waned and, at times, spontaneously resolved, whereas metastases from CIN-high

cells involved multiple organs and progressed rapidly, leading to death. Similar results were obtained after injection of lung adenocarcinoma H2030 cells (Fig. 2c–e and Extended Data Fig. 3a–c). Overexpression of the spindle assembly checkpoint protein MAD2 in MCAK-expressing cells partially rescued chromosome missegregation¹⁵ and correspondingly augmented metastasis (Fig. 2c and Extended Data Fig. 3f).

We performed orthotopic injections of MDA-MB-231 or 4T1 cells into the mammary fat pads of athymic or immune-competent BALB/c mice, respectively, followed by surgical excision of the primary tumour. Suppression of CIN had no effect on the efficiency of primary tumour implantation, and even enhanced primary tumour growth in the 4T1 model. However, in both models, suppression of CIN significantly reduced spontaneous metastasis and prolonged survival (Extended Data Fig. 3d, e).

We then assessed chromosome missegregation in the injected cells and in cells derived from primary tumours and metastatic colonies (Fig. 3a). We performed this analysis using MDA-MB-231 cells and two metastasis-competent xenografts (PDX) derived from patients with oestrogen receptor-positive (ER+) and triple-negative breast cancer (TNBC). Regardless of the CIN status of the injected cells, the majority of metastases were enriched for higher rates of chromosome missegregation, whereas cells derived from most primary tumours had significantly lower rates of CIN (Fig. 3b–d). For instance, when CIN-high cells (Fig. 3d, dnMCAK, blue bars) were injected into the mammary fat pad, chromosome missegregation rates decreased in the primary tumours (green bars) before increasing once more in metastases spontaneously arising within the same animal (orange bars).

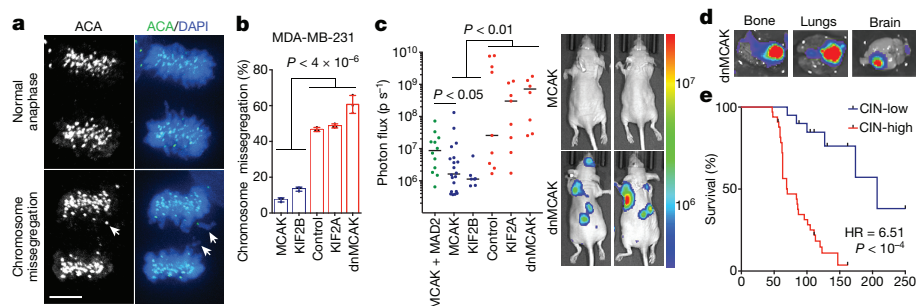


Figure 2 | CIN is a driver of metastasis. **a**, Anaphase cells stained for centromeres (ACA) and DNA (DAPI); scale bar, 5 μ m. **b**, Chromosome missegregation in MDA-MB-231 cells expressing kinesin-13 proteins. Bars represent mean \pm s.d., $n = 150$ cells. **c**, Whole animal bioluminescence (BLI) seven weeks after intracardiac injection of MDA-MB-231 cells. Left, bars represent the median and data points represent individual mice; $n = 12$ (MCAK + MAD2), 20 (MCAK), 7 (KIF2B), 9 (control), 9 (KIF2A), 8 (dnMCAK) mice. Right, representative images; colour scale shows

photon flux. **d**, Ex vivo BLI of organs with metastases from MDA-MB-231 cells expressing dnMCAK; colour scale as in **c**. **e**, Disease-specific survival of mice injected with CIN-high ($n = 33$) or CIN-low ($n = 20$) MDA-MB-231 cells. Significance tested using two-sided t -test (**b**), two-sided Mann–Whitney test (**c**), and two-sided log-rank test (**e**); $n = 3$ (**a**, **b**) and 5 (**d**) independent experiments. Throughout the paper, pairwise comparisons between individual CIN-low and CIN-high conditions are smaller than the stated P value. HR, hazard ratio.

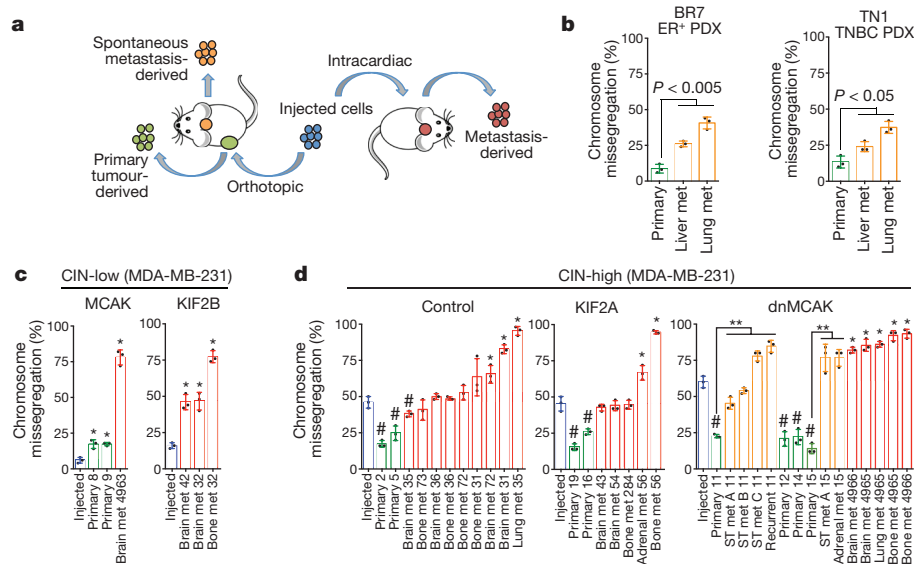


Figure 3 | Opposing roles for CIN in primary tumours and metastases. **a**, Experimental schema. **b–d**, Chromosome missegregation in injected cells (blue), cells derived from primary tumours (green), spontaneous metastases (met) (orange) or metastases arising after intracardiac injection (red). ST, soft tissue. Data shown as mean \pm s.d., $n = 150$ cells, three independent experiments; * $P < 0.05$, # $P < 0.05$, ** $P < 0.05$ denote samples with higher or lower chromosome missegregation than the injected lines, and spontaneous metastases with higher missegregation than matched primary tumours, respectively. Significance tested using two-sided t -test (**b–d**).

CIN enriches for mesenchymal traits

Bulk RNA sequencing (RNA-seq) identified 1,584 genes that were differentially expressed between CIN-low and CIN-high MDA-MB-231 cells. Principal component analysis and unsupervised clustering accurately separated samples according to their CIN status. Metastasis-related and epithelial-to-mesenchymal transition (EMT) gene sets were relatively enriched in CIN-high cells. The top 23 differentially expressed genes in CIN-high cells (referred to as CIN signature) predicted distant metastasis-free survival (DMFS) in a meta-analysis¹⁶ as well as a validation cohort¹⁷ of patients with breast cancer, irrespective of tumour subtype, grade, or lymph node status (Extended Data Figs 4, 5).

RNA-seq of primary tumour-derived and metastasis-derived cells revealed pathways that were shared among metastases and CIN-high cells. However, metastases contained a large number of differentially upregulated EMT and inflammation-related genes that were disproportionately clustered on chromosome 1, signifying chromosome 1-specific selection. Karyotype analysis revealed that the injected cell lines and most metastases had three copies of chromosome 1, whereas primary tumours consistently had two copies. Thus chromosome 1 loss is a recurrent event during primary tumour growth in this model (Extended Data Figs 4c–f, 5a–e).

We then performed single-cell RNA-seq (scRNA-seq) on three MDA-MB-231 cell lines—two CIN-low (KIF2B and MCAK) and one CIN-high (dnMCAK)—comprising a total of 6,821 cells. Clustering of single cells using EMT genes successfully classified most cells according to their CIN-status and revealed a fraction of cells (primarily CIN-high)

that was highly enriched in mesenchymal markers (Fig. 4a). Unsupervised graph-based clustering, based on all expressed genes, identified 12 phenotypically distinct subpopulations. One subpopulation was defined by increased expression of genes involved in EMT and metastasis (referred to as subpopulation ‘M’) and was concomitantly enriched for CIN signature genes. This subpopulation comprised 45% of dnMCAK expressing cells compared to 6% of CIN-low cells (Fig. 4b and Extended Data Fig. 6a, b).

In agreement with the scRNA-seq data, CIN-high cells exhibited increased migratory and invasive behaviour *in vitro*, and displayed evidence of actin cytoskeletal reorganization, diffuse vimentin staining, and increased cytoplasmic and nuclear localization of β -catenin (Extended Data Figs 6c, d, 7a–d). As expected, MAD2 overexpression rescued invasion and migration of MCAK-expressing cells. Furthermore, the ability of KIF2B or MCAK overexpression to suppress invasion *in vitro* was dependent on the cell cycle, as the addition of thymidine after transient transfection of either protein abrogated this phenotype (Extended Data Figs 6f, 7e, f and Supplementary Fig. 2).

CIN generates cytosolic DNA

To better define CIN-responsive pathways, we performed a gene–gene Pearson correlation analysis using scRNA-seq data and identified two large gene modules: module 1 was characterized by proliferative and metabolic pathways, whereas module 2 comprised EMT and inflammation gene sets (Fig. 5a). There was a strong positive correlation between inflammation-related, CIN signature, and EMT genes in the scRNA-seq and bulk RNA-seq data (Figs 4b, 5b and Extended Data Fig. 4b, c).

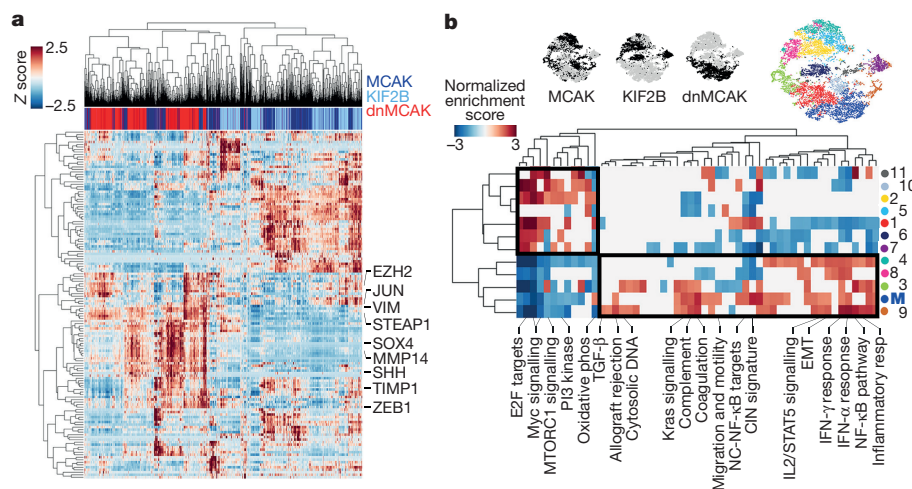
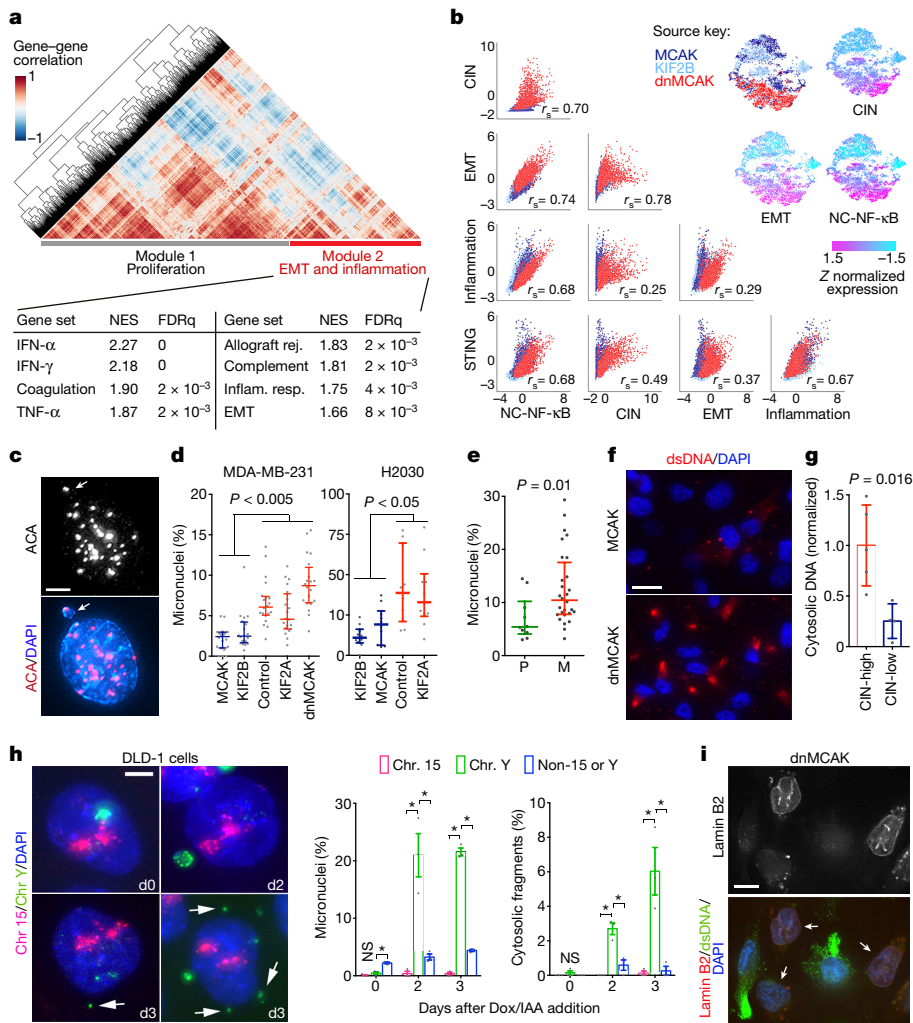


Figure 4 | CIN enriches for mesenchymal cell traits. **a**, Heat map showing expression of EMT genes in 6,821 MDA-MB-231 cells expressing MCAK, KIF2B, and dnMCAK. **b**, Top, t-stochastic neighbour embedding (tSNE) projection of all cells in **a**. Bottom, heat map of normalized enrichment score (NES) for pathways with an FDR $q < 0.05$ based on gene set enrichment analysis using one-sided weighted Smirnov–Kolmogorov test on subpopulations identified using Phenograph. Boxes outline populations with mutually exclusive transcriptional profiles.

**Figure 5 | CIN generates cytosolic DNA.**

a, Heat map showing gene-gene correlations in 6,821 cells and HALLMARK gene sets significantly enriched in module 2, one-sided weighted Smirnov-Kolmogorov test. **b**, Top right, normalized expression of key gene signatures in 6,821 MDA-MB-231 cells expressing MCAK, KIF2B, and dnMCAK. Bottom left, correlation plots for key gene signatures. **c**, A primary nucleus and a micronucleus stained for centromeres and DNA; scale bar 5 μ m. **d, e**, Percentage of micronuclei cells expressing various kinesin-13 proteins (**d**) or in cells derived from 10 primary tumours and 28 metastases (**e**). $n = 20$ high-power fields per sample (**d**) or average values derived from 10 high-power fields per sample (**e**). **f**, MDA-MB-231 cells expressing MCAK and dnMCAK stained using DAPI and anti-dsDNA antibody; scale bar, 20 μ m. **g**, Cytosolic-to-nuclear DNA ratios in CIN-high ($n = 5$) and CIN-low ($n = 4$) MDA-MB-231 ($n = 5$) and H2030 ($n = 4$) cells. **h**, Left, DLD-1 cells stained for DNA and hybridized to chromosome-specific FISH probes; scale bar, 5 μ m. Right, percentage of cells containing micronuclei or small cytosolic DNA fragments ($n > 500$ cells per condition). **i**, MDA-MB-231 cells stained using DAPI, anti-dsDNA antibody, or mCherry-lamin B2 (arrow); scale bar, 10 μ m. Data shown as median \pm interquartile range (**d, e**), mean \pm s.d. (**g**), mean \pm s.e.m. (**h**). $n = 6$ (**c, d**), 3 (**e, f, h**), 1 (**g**) and 2 (**i**) independent experiments. Significance tested using two-sided Mann-Whitney test (**d, e**) or two-sided t -test (**g, h**); $*P < 0.05$.

The induction of inflammatory pathways in response to CIN was unexpected and was reminiscent of a viral infection. We investigated whether CIN might introduce genomic DNA into the cytosol, thereby eliciting cellular responses normally reserved for anti-viral immunity^{18,19}. The exposure of genomic DNA to the cytosol can result from either primary nuclear or micronuclear envelope ruptures^{20–24}. We performed live-cell imaging using a GFP reporter with a nuclear localization signal (NLS-GFP)²⁵ and found no correlation between CIN and the frequency of NLS-GFP leakage into the cytosol in unconfined conditions. There was even a trend towards more efficient primary nucleus repair in CIN-high cells. CIN-high nuclei ruptured more frequently only during confined migration, and this was primarily attributed to their increased ability to go through a larger number of small constrictions (Extended Data Fig. 7g–j) that mimic confined migration during metastasis²⁵.

Instead, CIN-high cells and those derived from metastases exhibited a higher preponderance of micronuclei than did CIN-low or primary tumour-derived cells, respectively (Fig. 5c–e and Extended Data Fig. 8a–c). To test whether the presence of rupture-prone²⁰ micronuclei correlated with increased cytosolic DNA, we stained cells using two different anti-dsDNA antibodies after selective plasma membrane permeabilization and found increased cytosolic dsDNA and single-stranded DNA (ssDNA) in CIN-high cells. The dsDNA signal, which was distinct from mitochondrial staining, disappeared after treatment with double-strand-specific—but not single-strand-specific—nuclease and after overexpression of DNASE2, confirming the specificity of these antibodies (Fig. 5f and Extended Data Fig. 8d–h). Quantification of dsDNA levels after subcellular fractionation revealed a fourfold reduction in

cytosolic DNA in CIN-low cells compared to CIN-high cells (Fig. 5g). Whole-genome sequencing of subcellular fractions at 30 \times coverage confirmed the genomic origin of cytosolic DNA (data not shown).

To determine whether missegregated chromosomes provide a source of cytosolic DNA, we used an inducible Y-chromosome-specific missegregation system established in chromosomally stable DLD-1 colorectal cancer cells²⁶. Whole-chromosome fluorescence *in situ* hybridization (FISH) probes targeting the Y chromosome or an independent autosome (chromosome 15) revealed selective incorporation of the Y chromosome into micronuclei two days after chromosome missegregation induced by doxycycline and auxin (Dox/IAA) treatment. Notably, Y-chromosome-specific fragments were found dispersed within the cytosol 2–3 days after Dox/IAA addition, whereas the control autosome remained confined to the nucleus (Fig. 5h), demonstrating that cytosolic DNA is generated from chromosomes undergoing high rates of missegregation.

Suppression of micronuclear envelope rupture by mCherry-lamin B2 overexpression²⁰ reduced cytosolic dsDNA staining without influencing chromosome segregation errors. Accordingly, such overexpression reduced metastasis after intracardiac or tail vein injection of MDA-MB-231 cells (Fig. 5i and Extended Data Fig. 3g, h).

Metastasis from cytosolic DNA response

In chromosomally stable cells, cytosolic dsDNA is scarce and is sensed by the cGAS–STING pathway¹⁹, leading to induction of type I interferon stimulated genes (ISGs)^{22,23,27}. Indeed, induced missegregation of the Y chromosome led to the upregulation of OAS2, an ISG, and increased interferon- β production by DLD-1 cells

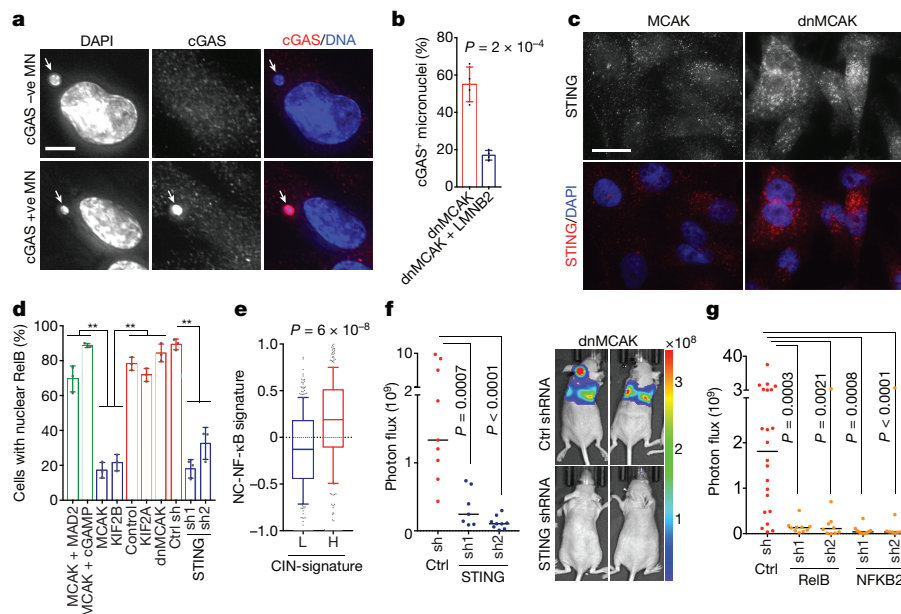


Figure 6 | Metastasis from a cytosolic DNA response. **a**, MDA-MB-231 cells stained for DNA and cGAS; scale bar, 5 μ m. **b**, Percentage of micronuclei with cGAS localization (cGAS⁺), $n = 200$ cells. **c**, Cells stained for DNA and STING; scale bar, 20 μ m. **d**, Percentage of MDA-MB-231 cells with nuclear RELB, $n = 150$ cells. **e**, Average z-normalized expression of CIN-responsive noncanonical NF- κ B target genes in breast cancer patients with low (<30th percentile, $n = 330$) or high (>30th percentile, $n = 332$) CIN gene expression signature. **f**, **g**, Photon flux ($p s^{-1}$) of whole animals after intracardiac injection with

MDA-MB-231 cells expressing control shRNA, STING shRNA (**f**), RELB shRNA (**g**) or NFKB2 shRNA (**g**); $n = 9$, 7, and 9 mice for the control, STING shRNA1, and STING shRNA2 groups, respectively (**f**); $n = 22$, 10, 10, 10, and 9 mice for the control, RELB shRNA1, RELB shRNA2, NFKB2 shRNA1, and NFKB2 shRNA2 groups, respectively (**g**). Data shown as mean \pm s.d. (**b**, **d**), median \pm interquartile range with bars spanning 10th–90th percentiles (**e**), median (**f**, **g**); significance tested using two-sided t -test (**b**, **d**), two-sided Mann–Whitney test (**e**–**g**); $n = 4$ (**a**, **b**) and 3 (**c**, **d**) independent experiments.

(Extended Data Fig. 9f, g). It is unclear how chromosomally unstable cancer cells cope with the constant presence of cytosolic DNA. We found notable localization of cGAS to approximately half of all micronuclei, as previously observed^{22,23,27}. Impeding micronuclear rupture through lamin B2 overexpression²⁰ significantly diminished the relative fraction of cGAS⁺ micronuclei (Fig. 6a, b). Furthermore, CIN-high cells exhibited increased levels and perinuclear localization of STING, congruent with pathway activation (Fig. 6c).

Notably, there was no evidence for robust activation of downstream canonical NF- κ B or type I interferon signalling in CIN-high cells, as evidenced by the lack of a significant increase in p65 or IRF3 phosphorylation, absence of p65 or IRF3 nuclear translocation, undetectable levels of interferon- β , and failure to induce ISGs (Extended Data Figs 8i, j, 9), in line with previous observations^{28–30}.

Cytosolic DNA, however, can activate the noncanonical NF- κ B pathway in a STING-dependent and TBK1-independent manner¹⁸. We found evidence for noncanonical NF- κ B activation in CIN-high cells, as revealed by lower levels of the precursor protein p100, a trend towards higher ratios of p52 and phosphorylated p100 relative to total p100, and reduced levels of the noncanonical NF- κ B pathway inhibitor TRAF2³¹ (Extended Data Fig. 8i, j). Given the subtle differences seen at the protein level, we assessed the nuclear localization of RELB, the binding partner of p52, and observed increased nuclear localization in CIN-high cells. This was often accompanied by cytosolic staining, indicative of chronic pathway activation. STING depletion reduced nuclear localization of RELB and led to downregulation of EMT and inflammatory pathways, whereas the addition of cGAMP or overexpression of MAD2 increased nuclear RELB in MCAK-expressing cells (Fig. 6d, Extended Data Figs 4e, 9c–e).

Bulk RNA-seq data identified a number of noncanonical NF- κ B target genes that were upregulated in response to CIN (CIN-responsive NC-NF- κ B genes). There was a robust correlation between the CIN signature, STING, and the CIN-responsive NC-NF- κ B genes in scRNA-seq data, in contrast to a weaker correlation between CIN and

type I interferon targets (Fig. 5b and Extended Data Fig. 6e). Similarly, RNA-seq data from primary breast cancer in the TCGA database demonstrated increased expression of CIN-responsive NC-NF- κ B genes in tumours with higher levels of CIN signature gene products (Fig. 6e), and higher expression of key regulators of the noncanonical NF- κ B pathway or its CIN-responsive target genes was associated with shorter DMFS and disease-free survival in breast and lung cancers. Conversely, increased expression of canonical NF- κ B or type I interferon regulatory factors was associated with an improved prognosis (Extended Data Fig. 10a).

cGAS activation by cancer cells has been invoked in brain metastasis through a tumour cell-non-autonomous mechanism³². We found that STING and downstream noncanonical NF- κ B activity mediate metastasis in a tumour cell-autonomous fashion, as evidenced by reduction in metastatic dissemination, lifespan extension, and reduction in *in vitro* and *in vivo* invasion of CIN-high cells depleted of STING, RELB, or p100 (encoded by *NFKB2*). Conversely, the addition of cGAMP increased invasion and migration of CIN-low cells (Fig. 6f, g and Extended Data Figs 6f, 9j, 10b–d). These findings are in line with reported roles for the noncanonical NF- κ B pathway in EMT, cellular invasion, and metastasis^{33–35}. The benefits of the noncanonical pathway may justify the scarcity of inactivating mutations in cGAS and STING among breast and lung cancers (Extended Data Fig. 10e).

Discussion

Our work reveals an unexpected link between CIN, chronic activation of cytosolic DNA sensing pathways, and metastasis. In addition to fuelling karyotypic heterogeneity that serves as a substrate for natural selection, ongoing chromosome missegregation is required to replenish cytosolic DNA pools and to maintain cells in a pro-metastatic state. Consequently, suppression of CIN reduces metastasis even in highly aneuploid cells. The repercussions of STING activation are context-dependent and range from senescence to tumorigenesis^{21–23,27,28,30}.

Given that chromosomally unstable cells are awash with cytosolic DNA, our results raise the possibility that by suppressing downstream type I interferon signalling³⁰ and instead upregulating the alternative NF- κ B pathway, such cells have substituted a lethal epithelial response to inflammation with that of myeloid-derived cells^{36,37}, thereby engaging in some form of immune mimicry. Restoration of normal responses to inflammation would constitute a viable therapeutic strategy to target chromosomally unstable cells.

The emergence, and subsequent tolerance, of CIN represents an important bottleneck during tumour evolution^{38–40}. We found that CIN induces a transcriptional shift from a proliferative and highly metabolic state, ideally suited for primary tumour growth, to a mesenchymal state associated with upregulation of inflammatory pathways (Figs 4b, 5a). These two mutually exclusive states, which were recently observed in a pan-cancer genomic analysis of metastatic tumours⁴¹, are likely to account for the reversibility in chromosome missegregation rates seen in primary tumours and metastases, and provide an explanation for the negative effect of aneuploidy during early tumorigenesis^{38,40,42}. These findings also lead us to suggest that CIN drives the subset of human metastases characterized by EMT and inflammation⁴¹.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 April; accepted 6 December 2017.

Published online 17 January 2018.

- Jamal-Hanjani, M. *et al.* Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).
- Turajlic, S. & Swanton, C. Metastasis as an evolutionary process. *Science* **352**, 169–175 (2016).
- Thompson, S. L. & Compton, D. A. Examining the link between chromosomal instability and aneuploidy in human cells. *J. Cell Biol.* **180**, 665–672 (2008).
- Cimini, D. *et al.* Merotelic kinetochore orientation is a major mechanism of aneuploidy in mitotic mammalian tissue cells. *J. Cell Biol.* **153**, 517–527 (2001).
- Bakhom, S. F. *et al.* The mitotic origin of chromosomal instability. *Curr. Biol.* **24**, R148–R149 (2014).
- Bakhom, S. F. *et al.* Numerical chromosomal instability mediates susceptibility to radiation treatment. *Nat. Commun.* **6**, 5990 (2015).
- Bakhom, S. F., Thompson, S. L., Manning, A. L. & Compton, D. A. Genome stability is ensured by temporal control of kinetochore–microtubule dynamics. *Nat. Cell Biol.* **11**, 27–35 (2009).
- Burrell, R. A. *et al.* Replication stress links structural and numerical cancer chromosomal instability. *Nature* **494**, 492–496 (2013).
- Brastianos, P. K. *et al.* Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5**, 1164–1177 (2015).
- Mitelman, F., Johansson, B. & Mertens, F. *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer* <https://cgap.nci.nih.gov/Chromosomes/Mitelman> (2017).
- Chung, C. H. *et al.* Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer Cell* **5**, 489–500 (2004).
- Moore, A. T. *et al.* MCAK associates with the tips of polymerizing microtubules. *J. Cell Biol.* **169**, 391–397 (2005).
- Ems-McClung, S. C. & Walczak, C. E. Kinesin-13s in mitosis: key players in the spatial and temporal organization of spindle microtubules. *Semin. Cell Dev. Biol.* **21**, 276–282 (2010).
- Waterman-Storer, C. M., Worthylyke, R. A., Liu, B. P., Burrage, K. & Salmon, E. D. Microtubule growth activates Rac1 to promote lamellipodial protrusion in fibroblasts. *Nat. Cell Biol.* **1**, 45–50 (1999).
- Sotillo, R. *et al.* Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. *Cancer Cell* **11**, 9–23 (2007).
- Györfy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res. Treat.* **123**, 725–731 (2010).
- Hatzis, C. *et al.* A genomic predictor of response and survival following taxane–anthracycline chemotherapy for invasive breast cancer. *J. Am. Med. Assoc.* **305**, 1873–1881 (2011).
- Abe, T. & Barber, G. N. Cytosolic-DNA-mediated, STING-dependent proinflammatory gene induction necessitates canonical NF- κ B activation through TBK1. *J. Virol.* **88**, 5328–5341 (2014).
- Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP–AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786–791 (2013).
- Hatch, E. M., Fischer, A. H., Deerinck, T. J. & Hetzer, M. W. Catastrophic nuclear envelope collapse in cancer cell micronuclei. *Cell* **154**, 47–60 (2013).
- Ahn, J. *et al.* Inflammation-driven carcinogenesis mediated through STING. *Nat. Commun.* **5**, 5166 (2014).
- Harding, S. M. *et al.* Mitotic progression following DNA damage enables pattern recognition within micronuclei. *Nature* **548**, 466–470 (2017).
- Mackenzie, K. J. *et al.* cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **548**, 461–465 (2017).
- Lan, Y. Y., Londoño, D., Bouley, R., Rooney, M. S. & Hacohen, N. Dnase2a deficiency uncovers lysosomal clearance of damaged nuclear DNA via autophagy. *Cell Reports* **9**, 180–192 (2014).
- Denais, C. M. *et al.* Nuclear envelope rupture and repair during cancer cell migration. *Science* **352**, 353–358 (2016).
- Ly, P. *et al.* Selective Y centromere inactivation triggers chromosome shattering in micronuclei and repair by non-homologous end joining. *Nat. Cell Biol.* **19**, 68–75 (2017).
- Yang, H., Wang, H., Ren, J., Chen, Q. & Chen, Z. J. cGAS is essential for cellular senescence. *Proc. Natl Acad. Sci. USA* **114**, E4612–E4620 (2017).
- Stetson, D. B., Ko, J. S., Heidmann, T. & Medzhitov, R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* **134**, 587–598 (2008).
- Lau, L., Gray, E. E., Brunette, R. L. & Stetson, D. B. DNA tumor virus oncogenes antagonize the cGAS–STING DNA-sensing pathway. *Science* **350**, 568–571 (2015).
- Dou, Z. *et al.* Cytoplasmic chromatin triggers inflammation in senescence and cancer. *Nature* **550**, 402–406 (2017).
- Sun, S.-C. Controlling the fate of NIK: a central stage in noncanonical NF- κ B signaling. *Sci. Signal.* **3**, pe18 (2010).
- Chen, Q. *et al.* Carcinoma–astrocyte gap junctions promote brain metastasis by cGAMP transfer. *Nature* **533**, 493–498 (2016).
- Wang, X. *et al.* Oestrogen signalling inhibits invasive phenotype by repressing RelB and its target BCL2. *Nat. Cell Biol.* **9**, 470–478 (2007).
- Wang, J., Yi, S., Zhou, J., Zhang, Y. & Guo, F. The NF- κ B subunit RelB regulates the migration and invasion abilities and the radio-sensitivity of prostate cancer cells. *Int. J. Oncol.* **49**, 381–392 (2016).
- Yang, C. *et al.* Antagonism of inhibitor of apoptosis proteins increases bone metastasis via unexpected osteoclast activation. *Cancer Discov.* **3**, 212–223 (2013).
- Gulen, M. F. *et al.* Signalling strength determines proapoptotic functions of STING. *Nat. Commun.* **8**, 427 (2017).
- Kis-Toth, K., Szanto, A., Thai, T. H. & Tsokos, G. C. Cytosolic DNA-activated human dendritic cells are potent activators of the adaptive immune response. *J. Immunol.* **187**, 1222–1234 (2011).
- Sheltzer, J. M. *et al.* Single-chromosome gains commonly function as tumor suppressors. *Cancer Cell* **31**, 240–255 (2017).
- Weaver, B. A. A., Silk, A. D., Montagna, C., Verdier-Pinard, P. & Cleveland, D. W. Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer Cell* **11**, 25–36 (2007).
- Rowald, K. *et al.* Negative selection and chromosome instability induced by Mad2 overexpression delay breast cancer but facilitate oncogene-independent outgrowth. *Cell Reports* **15**, 2679–2691 (2016).
- Robinson, D. R. *et al.* Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297–303 (2017).
- Sheltzer, J. M. A transcriptional and metabolic signature of primary aneuploidy is present in chromosomally unstable cancer cells and informs clinical prognosis. *Cancer Res.* **73**, 6401–6412 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Massagué, R. Benezra, X. Cai, J. Leeman, M. Bakhom, B. Hopkins and D. Landau for feedback. Grant support: S.F.B. (DoD Breast Cancer Research Breakthrough Award (BCRBA) W81XWH-16-1-0315, Elsa Pardee Foundation, MSKCC CytoGenetics Core (P30-CA008748) and Core (P30-CA008748) grants; B.N. (NSF Graduate Research Fellowship DGE1257284); J.L. (NIH R01-HL082792, U54-CA210184, DoD BCRBA BC150580, NSF CBET-1254846); P.L. (NCI K99-CA218871); G.G. (DoD BCRBA W81XWH-16-1-0316); L.C.C. (NIH R35-CA197588, U54-CA210184, Breast Cancer Research Foundation, Gray Foundation Basser Initiative).

Author Contributions S.F.B. and L.C.C. conceived the project; S.F.B., B.N., J.-A.C. and R.K.S. performed animal experiments; V.K.R. derived PDX models; P.L. performed experiments using DLD-1 cells; S.F.B., B.N., J.A.C., R.K.S., M.D., S.V. and Q.L. performed immunostaining, immunoblotting, microscopy and qPCR; P.S. performed live-cell imaging; T.B.K.W., N.J.B. and N.M. analysed matched primary tumour–metastasis data; N.K.T., C.H.C. and S.F.B. analysed the HNSCC data; S.F.B. analysed breast cancer karyotype data; C.P. performed histological analysis; C.S., K.C. and G.N. performed cytogenetic analysis; A.M.L. analysed scRNAseq data; M.L. and S.F.B. analysed survival data; C.J.M. analysed bulk RNAseq data; M.I. analysed whole-genome sequence data; G.G., M.L., Q.L., J.H.H., O.E., C.H.C., N.Y.L., D.P., D.W.C., S.N.P., J.L., C.S. and L.C.C. assisted with data interpretation. All authors contributed to the writing and editing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to L.C.C. (lcantley@med.cornell.edu).

Reviewer Information Nature thanks N. Gekara, J. van Deursen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Genomic analysis of primary tumour–metastasis matched pairs. Whole-exome DNA sequence data from 79 brain metastases with matched primary tumour and normal tissue⁹ were downloaded from the database of Genotypes and Phenotypes (dbGAP) and processed as described⁴³ to derive allele-specific segmented DNA copy number data for each sample. The weighted genome instability index (wGII), which describes the proportion of the genome that was classified as aberrant relative to tumour ploidy, was determined as described⁸.

Mitelman database analysis. All available breast adenocarcinoma cases in the Mitelman database¹⁰ were analysed. The primary literature was reviewed to determine the source of the sample (primary tumour or metastasis). When clonal karyotype was reported as a range, the average value was used for that given clone. Karyotype aberrations included structural aberrations as well as numerical deviations from the overall karyotype of the clone.

Analysis of chromosome segregation in head and neck squamous cell carcinoma. We analysed primary tumour specimens from 60 patients with head and neck squamous cell carcinoma (HNSCC)¹¹. Haematoxylin and eosin-stained primary tumour samples of sufficient quality for high-resolution microscopy analysis were available for forty patients. Analysis was restricted to cells fixed while undergoing anaphase, as previously described^{44,45}. Chromosome missegregation was defined by the presence of haematoxylin staining between the remaining segregating chromosomes during anaphase and was reported as the percentage of cells undergoing anaphase with evidence of chromosome missegregation, as previously described⁴⁴. Clinical lymph node status was defined on the basis of clinical examination or radiographic evidence of lymph node tumour involvement¹¹.

Single-cell karyotyping. Cultures were treated with colcemid at a final concentration of $0.1 \mu\text{g ml}^{-1}$. Following 45 min incubation at 37°C , the cultures were trypsinized, resuspended in pre-warmed 0.075 M KCl , incubated for an additional 10 min at 37°C and fixed in methanol:acetic acid (3:1). The fixed cell suspension was then dropped onto slides, stained in $0.08 \mu\text{g ml}^{-1}$ DAPI in $2 \times \text{SSC}$ for 5 min and mounted in antifade solution (Vectashield, Vector Labs). Metaphase spreads were captured using a Nikon Eclipse E800 epifluorescence microscope equipped with GenASI Cytogenetic suite (Applied Spectral Imaging). For each sample a minimum of 20 inverted DAPI-stained metaphases were fully karyotyped and analysed according to the International System of Human Cytogenetic Nomenclature (ISCN) 2013.

FISH analysis. FISH analysis was performed on fixed cells prepared for single-cell karyotyping. Based on karyotype data, four chromosomes were selected to further evaluate numerical instability. Probes specific for centromere 3 (red), centromere 4 (orange) and centromere 9 (green) were purchased from Abbott-Vysis. The chromosome 6 centromeric probe was home-brew (PAC clone P308; labelled with green dUTP, MSKCC Molecular Cytogenetics Core Facility). Probe labelling, hybridization, post-hybridization washing, and fluorescence detection were performed according to standard laboratory procedures²⁶. For each probe, three normal peripheral blood samples (GM07535, GM06875 and GM00558), obtained from Coriell Institute, were also analysed to derive cut-off values (false positive).

RHOA and RAC1 pull-down assays. The activity of RHOA and RAC1 was determined using bead-based pull-down assay kits (Cytoskeleton, RHOA: BK036S, RAC1: BK035S). Cells were lysed on the tissue culture dish and rapidly snap frozen for storage until further processing. cGAMP was added for 18 h before lysis. In addition to His-tagged RHOA and RAC1, the positive and negative controls were total cell lysates supplemented with non-hydrolysable GTP or GDP, respectively.

Cell culture. Cell lines were purchased from the American Type Culture Collection (ATCC). Tumour (MDA-MB-231, 4T1, HEK293, and H2030), cells were cultured in DMEM or RPMI (4T1) supplemented with 10% FBS and 2 mM L-glutamine in the presence of penicillin (50 U ml^{-1}) and streptomycin ($50 \mu\text{g ml}^{-1}$). All cells tested negative for mycoplasma. Cell confluence was measured using IncuCyte live-cell analysis system (Essen Bioscience).

Immunofluorescence microscopy. Cell fixation and antibody staining were performed as previously described⁷. In brief, cells were fixed with ice-cold (-30°C) methanol for 15 min (when staining for centromeres, centrosomes, cGAS, Vimentin, β -actin, IRF3, or α -tubulin) or 4% paraformaldehyde (when staining for RELB, p65, STING, ssDNA, dsDNA, COXIV, or β -catenin). Subsequently, cells were permeabilized using 1% triton for 4 min. See Supplementary Table 1 for antibody information. For selective plasma membrane permeabilization used for cytosolic dsDNA and ssDNA staining, cells were treated with 0.02% saponin for 5 min after fixation. For single-stranded (Thermo Fisher Scientific, FEREN0321) and double stranded (Life Technologies, EN0771)-specific nuclease treatment, cells were also permeabilized with 0.02% saponin for 2 min and treated with either nuclease for 10 min before fixation using 4% paraformaldehyde. TBS-BSA was used as a blocking agent during antibody staining. DAPI was added together with secondary

antibodies. Cells were mounted with Prolong Diamond Antifade Mountant (Life Technologies, P36961). cGAMP (Invivogen, tlr-nacga23) was transfected into cells at a concentration of $4 \mu\text{g ml}^{-1}$ using lipofectamine 2000 that was added for 3–4 h and then replaced with regular serum-containing medium.

Immunoblotting. Cells were pelleted and lysed using RIPA buffer. Protein concentration was determined using BCA protein assay and 20–30 μg total protein was loaded in each lane. Proteins were separated by gradient SDS-PAGE and transferred to PVDF or nitrocellulose membranes. Full blots are shown in Supplementary Fig. 1. See Supplementary Table 2 for antibody information. Band intensities on immunoblots were obtained using ImageJ (<https://imagej.nih.gov/ij/>) or the LI-COR Odyssey software, normalized to loading control, and background was subtracted. Ratios were normalized to control cells. Interferon- β levels from conditioned medium were measured using the Human IFN beta Array 1-plex (Eve Technologies, HIFNB-01-31).

Y chromosome missegregation and FISH. Flp-In T-REx DLD-1 cells were engineered to express the TIR1 auxin-dependent plant E3 ligase, an auxin-inducible degron (AID)-tagged CENP-A modified at the endogenous allele (CENP-A^{AID/-}), and a doxycycline-inducible CENP-A^{C-H3} rescue gene integrated into the Flp-In locus as previously described²⁶. Cells (4.0×10^4) were seeded into 4-well chamber slides and treated with doxycycline (DOX, Sigma-Aldrich) and indole-3-acetic acid (IAA, Sigma-Aldrich) for up to 3 days to induce Y chromosome missegregation and micronuclei. Slides were washed in PBS, fixed in 3:1 methanol:acetic acid for 15 min at room temperature, and dehydrated with 80% ethanol. Chromosome paint FISH probes targeting chromosomes Y and 15 (MetaSystems) were mixed at equal ratios, applied to cells, sealed with a coverslip, and co-denatured at 75°C for 2 min followed by overnight hybridization at 37°C in a humidified chamber. Slides were washed in $0.4 \times$ saline–sodium citrate (SSC) buffer for 2 min at 72°C , followed by a 30 s wash in $2 \times \text{SSC}$, 0.05% Tween-20 buffer at room temperature. Cells were counterstained with DAPI and captured on a DeltaVision Elite (GE Healthcare) microscope system at $60 \times$ magnification ($25 \times 0.2 \mu\text{m}$ z-stacks) followed by image deconvolution and maximum intensity quick projection.

Knockdown and overexpression constructs. Luciferase expression was achieved using pLVX plasmid (expressing tdTomato) and cells stably expressing luciferase were selected using hygromycin and sorted for tdTomato expression. Kinesin-13 family protein expression was achieved using plasmid (pEGFP) transfection or lentiviral (pLenti-GIII-CMV-GFP-2A-Puro) expression where cells were selected using G418 (0.5 mg ml^{-1}) or puromycin ($5 \mu\text{g ml}^{-1}$), respectively. DNASE2 overexpression was achieved using a pLenti-GIII-CMV-RFP-2A-Puro plasmid with puromycin used for selection. Plasmids containing kinesin-13 proteins or lamin B2 (pQCXIB-mCherry-lmnB2) constructs were offered by the Compton and Hetzer Laboratories, respectively. Blasticidin was used to select for *lmnB2*-expressing cells at $10 \mu\text{g ml}^{-1}$. All other plasmids were purchased from Applied Biological Materials (<https://www.abmgood.com/>). Stable knockdown of STING, NFKB2, RELB, and cGAS were achieved using shRNAs in pRRL (SGEP or SGEN) plasmids obtained from the MSKCC RNA Interference Core. Two to four distinct shRNA hairpins were screened per target. Targeted shRNA sequences are listed in Supplementary Table 3. To visualize primary nuclear rupture, cells were stably modified with a retroviral construct expressing both NLS-GFP²⁵ and H2B-Tdtomato ($3 \times \text{NLS-GFP-P2A-H2B-Tdtomato-IRES-puromycin}$). Cells were cultured for 24 h after viral transduction before selection with $1 \mu\text{g ml}^{-1}$ puromycin and subsequently sorted to select for NLS-GFP and H2B-Tdtomato expression.

Cell migration in microfluidic devices. Microfluidic migration devices with precisely defined constrictions were prepared as described previously^{25,46}. Devices were coated with $50 \mu\text{g ml}^{-1}$ type-I rat tail collagen (BD Biosciences) in 0.02 M acetic acid overnight at 4°C . Approximately 80,000 cells were seeded (in DMEM supplemented with 10% FBS and 1% PenStrep) per migration chamber. Devices were placed in a tissue-culture incubator (37°C) for 5–6 h to allow the cells to adhere. Subsequently, the medium was changed to phenol-red free Leibovitz L15 medium supplemented with 10% FBS and 1% PenStrep before the device was mounted on an inverted microscope (Zeiss Observer Z1) equipped with a temperature-controlled stage (37°C) for live-cell imaging. The medium reservoirs of the device were covered with glass coverslips to minimize evaporation during live-cell imaging. Cells were imaged for 14–16 h at 10-min intervals with a CCD camera (Photometrics Coolsnap KINO) using a Zeiss $20 \times \text{NA } 0.8$ air objective. Acquired image sequences were analysed for nuclear rupture frequency, duration, and transit time of cells through $1 \times 5\text{-}\mu\text{m}^2$, $2 \times 5\text{-}\mu\text{m}^2$, and $15 \times 5\text{-}\mu\text{m}^2$ constrictions using Zen software (Zeiss) and a custom-written MATLAB 2016a script for automated image analysis.

Animal studies. Animal experiments were performed in accordance with protocols approved by the Weill Cornell Medicine Institutional Animal Care and Use Committee. For disease-specific survival in MDA-MB-231 experiments, power analysis indicated that ten mice per group would be sufficient to detect a difference

at relative hazard ratios of <0.2 or >5 with 80% power and 95% confidence, given a median disease-specific survival of 3 months in the control group and a total follow up period of 250 days. For the 4T1 experiments, power analysis indicated that ten mice per group would be sufficient to detect a difference at relative hazard ratios of <0.25 or >4.0 with 80% power and 95% confidence, given a median survival of 58 days in the control group and a total follow up period of 180 days. There was no need to randomize animals. Investigators were not blinded to group allocation. Intracardiac injection was performed as previously described³². In brief, cells were trypsinized and washed with PBS and 1×10^5 cells (in 100 μ l PBS) were injected into the left cardiac ventricle of female athymic 6–7-week-old athymic nude (*nu/nu*) mice (Jackson Laboratory strain 002019). Cells (2×10^5) were injected into the tail vein cohort of animals. Mice were then immediately injected with D-luciferin (150 mg kg⁻¹) and subjected to bioluminescence imaging (BLI) using tan IVIS Spectrum Xenogen instrument (Caliper Life Sciences) to ensure systemic dissemination of tumour cells. Metastatic burden was measured 5–7 weeks after injection using BLI. BLI images were analysed using Living Image Software v.2.50. Disease-specific survival endpoint was met when the mice died or met the criteria for euthanasia under the IACUC protocol and had radiographic evidence of metastatic disease. For orthotopic tumour implantation, 2.5×10^5 cells in 50 μ l PBS were mixed 1:1 with Matrigel (BD Biosciences) and injected into the fourth mammary fat pad. Only one tumour was implanted per animal. MDA-MB-231 primary tumours were surgically excised before they reached ~ 1.5 cm in the largest dimension (which was the maximum allowable under our IACUC protocol) and metastatic dissemination was assessed using BLI imaging at 1–3-week intervals for up to 30 weeks. The distant metastasis-free survival endpoint was met when BLI signal was seen outside the site of primary tumour transplantation. 4T1 tumours were excised 8–9 days after implantation. To derive short-term culture from primary tumours and metastases, anesthetized mice (isofluorane) were imaged then killed. *Ex vivo* BLI was subsequently performed on removed organs to define the precise location of the metastatic lesion. Primary tumours and metastases were subsequently mechanically dissociated and cultured in DMEM with selection medium (G418 or hygromycin) to select for tumour cells and exclude host cells. All subsequent assays (karyotyping, RNA-seq, immunofluorescence, and subcellular fractionation) were performed after a single passage from the primary sample. To assess chromosome missegregation from primary tumour-derived and metastasis-derived cells, we performed high-resolution immunofluorescence analysis on passage no. 1 cells, staining for DNA (DAPI) and centromeres (ACA). The presence of cells with DNA or centromere staining in the middle of the anaphase plate was taken as evidence of chromosome missegregation.

Patient-derived xenograft assays. Patient-derived xenograft (PDX) models of human metastatic breast cancers were generated by transplanting the freshly obtained surgically excised tumour specimens from patients who had given consent under the IRB approved protocol (MSKCC IRB #97-094) in female NOD/SCID/IL2Ry^{null} (NSG) mice (Jackson Laboratories strain 005557). All relevant ethical regulations were followed. The oestrogen receptor-positive PDX was derived from breast cancer metastatic to the bone. The TNBC PDX was established from an axillary lymph node metastasis from a patient with inflammatory breast cancer. PDXs were maintained for a maximum of three serial passages. In brief, freshly obtained tumour tissue specimens were either directly transplanted in the mammary fat-pad of the mice or minced into 1–2-mm pieces in serum-free MEM medium with nonessential amino acids (Cat. No. 41500018, Thermo Fisher Scientific) transduced with lentiviral vectors expressing either GFP-luciferase or pUltra-Chili-Luc plasmid (Addgene plasmid: 48688) followed by transplantation into mice. Typically PDX tumour growth became evident during the first 1–3 weeks after engrafting and tumours continued to grow for an additional 4–8 weeks. Primary tumour growth and metastases were followed using BLI or spectrum CT imaging. At the time of removal of primary tumours and metastases, we derived primary cell cultures directly from primary tumours as well as lung and liver metastases. In brief, 500 mg of fresh bulk tumour tissue was chopped into 1–2-mm³ pieces and incubated in Accutase (AT104; Innovative Cell Technologies) for cell detachment and separation over 1–2 h. The dissociated tissues were sieved through 100- μ m cell strainers and the cells were pelleted by centrifugation at 1,200 r.p.m. The pellets were washed and resuspended in the above MEM buffer with 3% FBS. Cells were analysed for chromosome missegregation after one passage.

RNA sequencing and analysis. Bulk RNA was extracted from cells using the QIAshredder (Qiagen, 79654) and the RNA extraction kit (Qiagen, 74106) and sequenced using HiSeq2500 or HiSeq4000 (Illumina). The quality of the raw FASTQ files were checked with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). For samples originating from mouse xenografts, FASTQ reads were classified as originating from either mouse (GRCm38) or human (GRCh38) genomes using xenome⁴⁷, and human-specific reads were used for mapping. Reads were mapped to human reference GRCh38 using STAR (v2.4.1d,

2-pass mode)⁴⁸. Gene expression was estimated using cufflinks (v2.2.1, default parameters) and HTSeq (v0.6.1)^{49,50}. Differential expression analyses were performed using DESeq2 (v1.14.1)⁵¹. Pre-ranked gene set enrichment analyses were performed on the DESeq2 log₂ fold changes⁵². Prior to any unsupervised analyses, expression counts were transformed using variance stabilizing transformation using the DESeq2 R package. Gene signatures used in the study are listed in Supplementary Table 5. Differentially expressed gene sets and their associated statistics are listed in Supplementary Table 6. To detect potential copy number changes, positional gene enrichment analysis (PGE) was performed on the upregulated and downregulated differentially expressed genes, separately ($\text{padj} \leq 0.1$)⁵³. Only significant regions with four or more genes and with $P \leq 0.01$ were kept for further analysis. Circos plots were made using the circize R package⁵⁴.

Reverse-transcriptase quantitative polymerase chain reaction. Cells were collected into trizol reagent (Thermo Fisher Scientific) and total RNA was extracted using 'PureLink RNA mini kit' (Thermo Fisher Scientific) according to the manufacturer's instructions. Total RNA (5 μ g) was used for RT-PCR using the RNA to cDNA Ecodym premix (oligo dT) cDNA synthesis kit (Clontech) according to the manufacturer's instructions. Resulting cDNA corresponding to 50 ng total RNA was used in each 20 μ l of quantitative real time PCR reaction. qRT-PCR was performed using SybrGreen master mix (Biorad) and the relative expression of each gene was calculated after normalizing to *ACTB* endogenous control and using the comparative $\Delta\Delta C_t$ method. A list of the primers used is in Supplementary Table 4.

Single-cell RNA sequencing. Cells were trypsinized and resuspended in PBS. Twenty-one microlitres of a cellular suspension at 400 cells per microlitre, $>95\%$ viability, were loaded onto the 10X Genomics Chromium platform to generate barcoded single-cell GEMs. scRNA-seq libraries were prepared according to 10X Genomics specifications (Single Cell 3' Reagent Kits User Guide PN-120233, 10X Genomics). GEM-reverse transcription (RT) (55°C for 2 h, 85°C for 5 min; held at 4°C) was performed in a C1000 Touch Thermal cycler with 96-Deep Well Reaction Module (Bio-Rad). After RT, GEMs were broken and the single-strand cDNA was cleaned up with DynaBeads MyOne Silane Beads (Thermo Fisher Scientific) and SPRIselect Reagent Kit (0.6 \times SPRI; Beckman Coulter). cDNA was amplified for 14 cycles using the C1000 Touch Thermal cycler with 96-Deep Well Reaction Module (98°C for 3 min; 98°C for 15 s, 67°C for 20 s, and 72°C for 1 min \times 14 cycles; 72°C for 1 min; held at 4°C). The quality of the cDNA was analysed using an Agilent Bioanalyzer 2100. The resulting cDNA was sheared to ~ 200 bp using a Covaris S220 instrument (Covaris) and cleaned using 0.6 \times SPRI beads. The products were end-repaired, 'A'-tailed and ligated to adaptors provided in the kit. A unique sample index for each library was introduced through 10 cycles of PCR amplification using the indexes provided in the kit (98°C for 45 s; 98°C for 20 s, 60°C for 30 s, and 72°C for 20 s \times 14 cycles; 72°C for 1 min; held at 4°C). After two SPRI cleanups, libraries were quantified using Qubit fluorometric quantification (Thermo Fisher Scientific) and the quality assessed on an Agilent Bioanalyzer 2100. Four libraries were pooled and clustered on a HiSeq2500 in rapid mode at 10 pM on a pair end read flow cell and sequenced for 98 cycles of R1, followed by 14 bp I7 Index (10X Barcode), 8 bp I5 Index (sample index) and 10 bp on R2 (UMI). Primary processing of sequencing images was done using Illumina's Real Time Analysis software. Demultiplexing and post processing was done using the 10X Genomics Cell Ranger pipeline per the manufacturer's recommendations. scRNA-seq data were processed from raw reads to a molecule count array using the Cell Ranger pipeline⁵⁵. Additionally, to minimize the effects of experimental artefacts on the analysis, data were pre-processed to filter out cells with low total molecule counts (library size), low complexity and high mitochondrial content, identified by a bimodal fit. The remaining cells were normalized by dividing the expression level of each gene in a cell by its total library size and then scaling by the median library size of all cells. After normalizing by library size, we performed principal component analysis to improve the robustness of the constructed Markov matrix generated when computing diffusion eigenvalues for imputation of dropout noise⁵⁶. We chose the number of principal components to retain approximately 80% of variance in the data and excluded the first principal component, which was highly correlated with library size. Imputation of both the normalized and unnormalized count matrix was performed using a Markov matrix raised to the power of 3 (power corresponds the approximate number of weighted nearest neighbours) and with a gene expression distribution computed according to 21 nearest neighbouring cells, as described⁵⁶. Our analysis was robust to imputation and we obtained similar results without imputed data (not shown). Subpopulations were identified using Phenograph⁵⁷ and genes differentially expressed in at least one subpopulation were identified by the Kruskal–Wallis rank statistic using a bootstrapping method for random downsampling of matched molecule and cell counts from each subpopulation. t-SNE was used to visualize subpopulation structure based on the first 20 principle components of the imputed count matrix, subsetted by the top

5,150 differentially expressed genes (FDR q of Kruskal–Wallis rank statistic <0.05). The mean expression of key gene signatures in population M versus other subpopulations was z -normalized and visualized using violin plots. All gene signatures are annotated in Supplementary Table 5. The correlation between gene signatures was computed using the Spearman rank correlation coefficient according to the mean expression of all genes per signature per cell. Ward's minimum variance method was applied to hierarchically cluster cells by their normalized expression of differentially expressed EMT genes.

Patient survival analysis. Genes used for survival analysis are listed in Supplementary Table 5. For the meta-analysis cohort, we used aggregate data from KMPLOT^{16,58} (<http://www.kmplot.com>) using only JetSet best probe set and auto-selection for best cutoff between the 25th and 75th percentiles. For the validation cohort in which DMFS data were available¹⁷, we used the z -normalized expression data for a data set and the median value was used as a cutoff. DMFS curves were compared using the log-rank test.

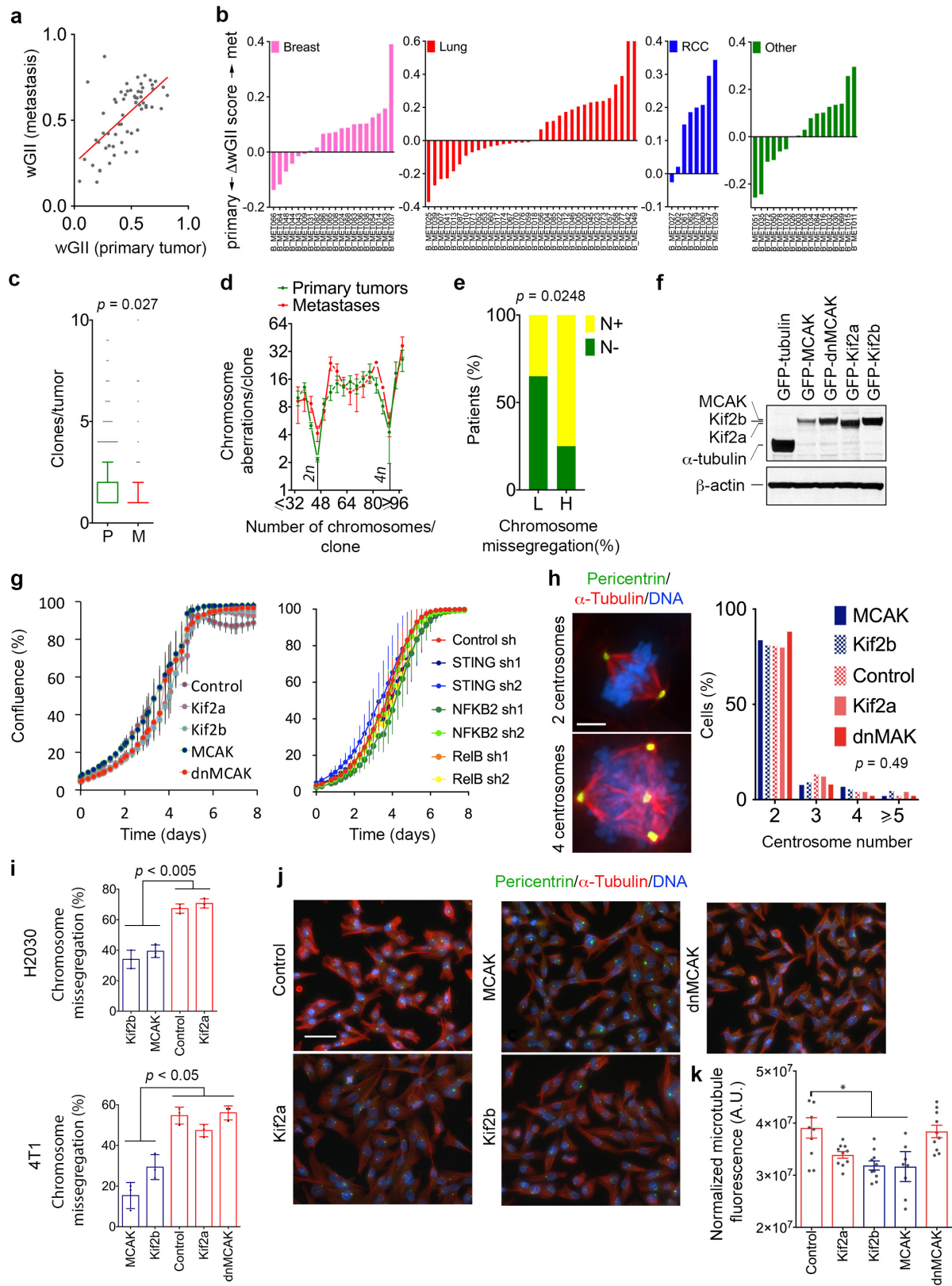
In vitro invasion and migration assays. For the invasion and migration/chemotaxis assays we used the CytoSelect cell invasion (CBA-110) and cell migration (CBA-100) kits, respectively. In brief, 3×10^5 cells were suspended in serum-free medium and placed on top of the membrane. Medium containing serum was placed at the bottom and cells that had invaded to the inferior surface of the collagen membrane were stained and counted 18–24 h later. For experiments involving transient transfection, cells were transfected, and thymidine (2 mM) was added 18 h later. Cells were plated on the membrane 3 days after transfection. For the chemotaxis assay, we used a colorimetric approach (OD 560 nm) for quantification. For the scratch assay, cells were treated with mitomycin C ($10 \mu\text{g ml}^{-1}$) for 1 h when they reached $>90\%$ confluence and then placed in DMEM containing 1% FBS. Wounds were applied using a p200 pipette tip and images of the wounds were taken immediately and at subsequent regular intervals. ImageJ was used for quantification of wound surface area.

Quantification of cytosolic DNA. Approximately 1×10^7 cells were lysed and the nuclear, cytosolic, and mitochondrial fractions were obtained using the mitochondrial isolation kit (Thermo Fisher Scientific, 89874). Protease inhibitors were not used to enable subsequent DNA purification. Mitochondria were purified by centrifugation at 12,000g to minimize their contamination in the cytosolic fraction. DNA was subsequently isolated from the nuclear, cytosolic, mitochondrial fractions using the Qiagen DNeasy blood and tissue kit (Qiagen, 69506) and dsDNA was quantified using Qubit 2.0 (Invitrogen) with Qubit dsDNA HS Reagent.

Code availability. All custom code, statistical analysis, and visualizations were performed in Python or R, and used Nextflow to manage some of the computational pipelines⁵⁹. Code for the RNA sequencing analysis is available online at: <https://github.com/murphycj/manuscripts/tree/master/BakhomEtAl2017>. The live-cell tracking MATLAB 2016a code can be found at <https://github.com/Lammerding/MATLAB-CellTracking>.

Data availability. Source data for Figs 1–3, 5, 6, and Extended Data Figs 1–3, 5–10 are provided with the paper. Single-cell RNA sequencing data (shown in Figs 4, 5 and Extended Data Fig. 6) have been deposited in the Sequence Read Archive under accession number SRP104750. Bulk RNA-seq data (shown in Extended Data Figs 4, 5) have been deposited in the Gene Expression Omnibus under accession number GSE98183.

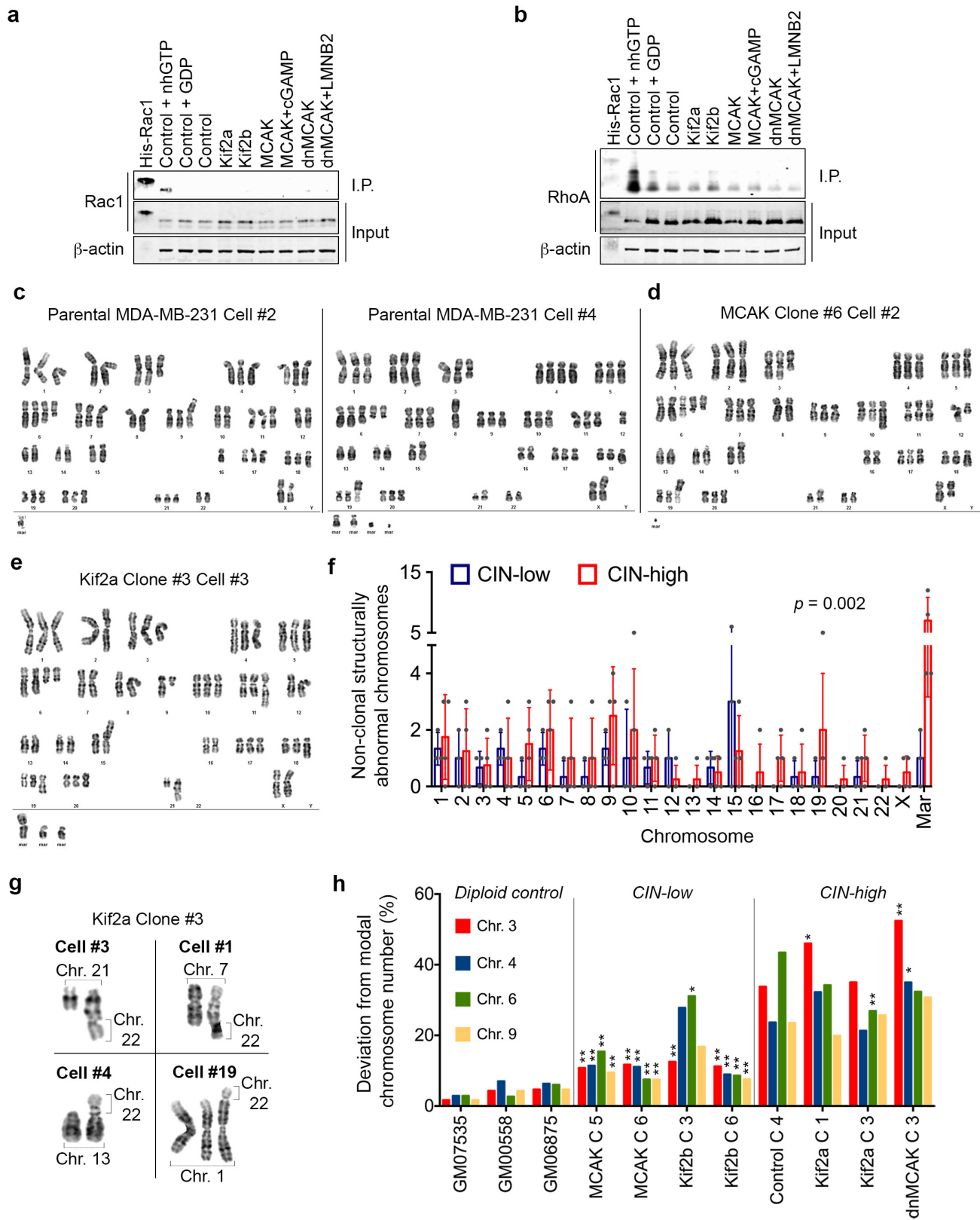
43. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351**, 1463–1469 (2016).
44. Bakhom, S. F., Danilova, O. V., Kaur, P., Levy, N. B. & Compton, D. A. Chromosomal instability substantiates poor prognosis in patients with diffuse large B-cell lymphoma. *Clin. Cancer Res.* **17**, 7704–7711 (2011).
45. Zaki, B. I., Suriawinata, A. A., Eastman, A. R., Garner, K. M. & Bakhom, S. F. Chromosomal instability portends superior response of rectal adenocarcinoma to chemoradiation therapy. *Cancer* **120**, 1733–1742 (2014).
46. Davidson, P. M., Denais, C., Bakshi, M. C. & Lammerding, J. Nuclear deformability constitutes a rate-limiting step during cell migration in 3-D environments. *Cell. Mol. Bioeng.* **7**, 293–306 (2014).
47. Conway, T. *et al.* Xenome—a tool for classifying reads from xenograft samples. *Bioinformatics* **28**, i172–i178 (2012).
48. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
49. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
50. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
51. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
52. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
53. De Preter, K., Barriot, R., Speleman, F., Vandesompele, J. & Moreau, Y. Positional gene enrichment analysis of gene sets for high-resolution identification of overrepresented chromosomal regions. *Nucleic Acids Res.* **36**, e43 (2008).
54. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
55. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
56. van Dijk, D. *et al.* MAGIC: A diffusion-based imputation method reveals gene–gene interactions in single-cell RNA-sequencing data. Preprint at <https://www.biorxiv.org/content/early/2017/02/25/111591> (2017).
57. Levine, J. H. *et al.* Data-driven phenotypic dissection of aml reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
58. Györfy, B., Surowiak, P., Budczies, J. & Lánczky, A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS ONE* **8**, e82241 (2013).
59. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35**, 316–319 (2017).



Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Generation of isogenic tumour models of CIN. **a**, wGII of brain metastases as a function of the wGII of the matched primary tumour. Red line represents linear regression, $n = 79$ patients. **b**, Differences in wGII between metastases and matched primary tumours. RCC, renal cell carcinoma; other includes melanoma, sarcoma, and ovarian, thyroid, and salivary gland cancers. **c**, Number of clones (based on single-cell karyotypes) in primary breast tumours (P; $n = 637$) or metastases (M; $n = 131$) found in the Mitelman database. Boxes represent median \pm interquartile range and bars span the 10th and 90th percentiles; significance tested using two-sided Mann–Whitney test. **d**, The number of chromosome aberrations per clone as a function of the total number of chromosomes in a given clone in samples derived from primary breast tumour clones ($n = 983$) and metastatic clones ($n = 186$); data shown as mean \pm s.d. **e**, Percentage of N $-$ or N $+$ patients as a function of chromosome missegregation frequency ($n = 20$ patients per condition); significance tested using two-sided Fisher's exact test. **f**, Immunoblots of cells expressing various GFP-tagged kinesin-13 proteins stained using anti-GFP antibody; β -actin used as a loading control, two independent experiments performed. **g**, Cellular confluence as a function

of time in MDA-MB-231 cells expressing various kinesin-13 proteins or dnMCAK-expressing cells depleted of components of the cytosolic DNA-sensing machinery or the noncanonical NF- κ B pathway. Data shown as mean \pm s.d., $n = 4$ independent experiments. **h**, Left, MCAK- and dnMCAK-expressing cells stained for microtubules (α -tubulin, DM1A), centrosomes (pericentrin) and DNA (DAPI). Scale bar, 5 μ m; two independent experiments performed. Right, frequency distribution of the number of pericentrin foci per cell. Significance tested using ANOVA. $n = 100$ cells per condition, two independent experiments performed. **i**, Chromosome missegregation in H2030 and 4T1 cells expressing kinesin-13 proteins. Data shown as mean \pm s.d., $n = 150$ cells, three independent experiments performed, significance tested using two-sided t -test. **j**, Cells expressing kinesin-13 proteins stained for microtubules (DM1A), centrosomes (pericentrin) and DNA (DAPI). Scale bar, 50 μ m, two independent experiments performed. **k**, Fluorescence normalized to cell count of MDA-MB0-231 cells expressing kinesin-13 proteins. Data shown as mean \pm s.e.m., $*P < 0.05$, two-sided t -test, $n = 10$ high-power fields encompassing 477–612 cells, two independent experiments performed.

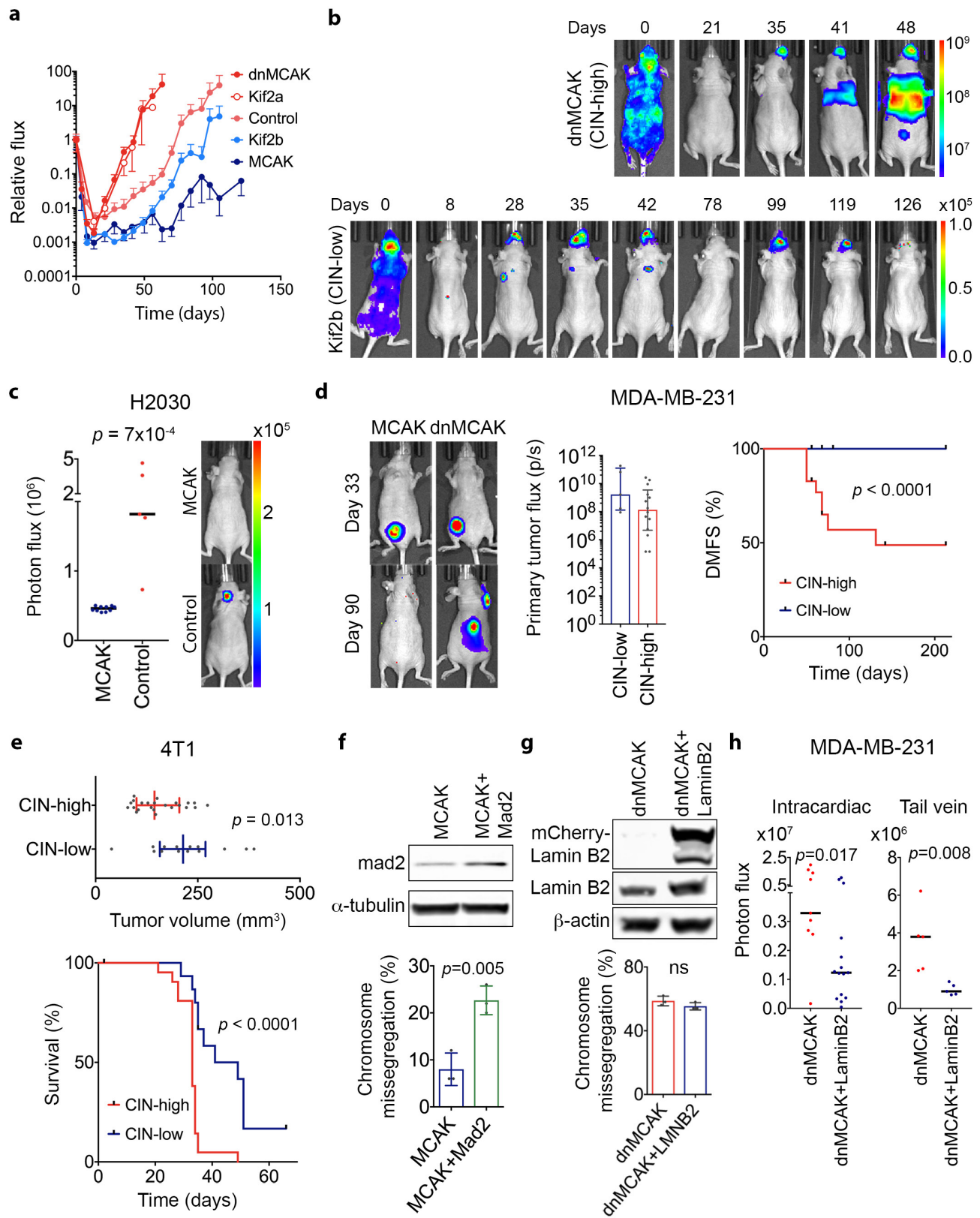


Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Karyotype analyses of human tumour cells.

a, b, Immunoblots showing total RAC1 (**a**) or RHOA (**b**) levels as well as RAC1 or RHOA that was pulled down using antibodies that were specific to the GTP-bound form of RAC1 (**a**) or RHOA (**b**). Positive and negative controls were total MDA-MB-231 cell lysates supplanted with non-hydrolysable GTP (nhGTP) and GDP, respectively. β -actin was used as a loading control, two independent experiments performed. **c–e**, Representative karyotypes (DAPI-banding) from parental MDA-MB-231 cells (**c**) or populations derived from single cells expressing MCAK (**d**) or KIF2A (**e**) that were allowed to divide for 30 days. **f**, The number of non-clonal (present in less than 25% of the cells in a single clone) structurally abnormal chromosomes in CIN-low or CIN- high MDA-MB-231 cells.

Mar, chromosomes so structurally abnormal that they could not be identified by conventional banding; data shown as mean \pm s.d., $n = 140$ cells from 7 clonal populations, significance tested using two-way ANOVA test. **g**, Examples taken from four distinct cells belonging to the same clonal population (derived from a single KIF2A-expressing cell) showing convergent translocations involving chromosome 22 with four other chromosomes. **h**, Deviation from modal chromosome number in single-cell-derived clones grown for 30 days. Four chromosomes were assayed for each clone using centromere-specific probes. * $P < 0.05$, ** $P < 0.005$ compared to control clone 4 by two-sided χ^2 -test, $n = 300$ cells per clone. Diploid controls were used to determine the false-positive rate of the centromeric probes.

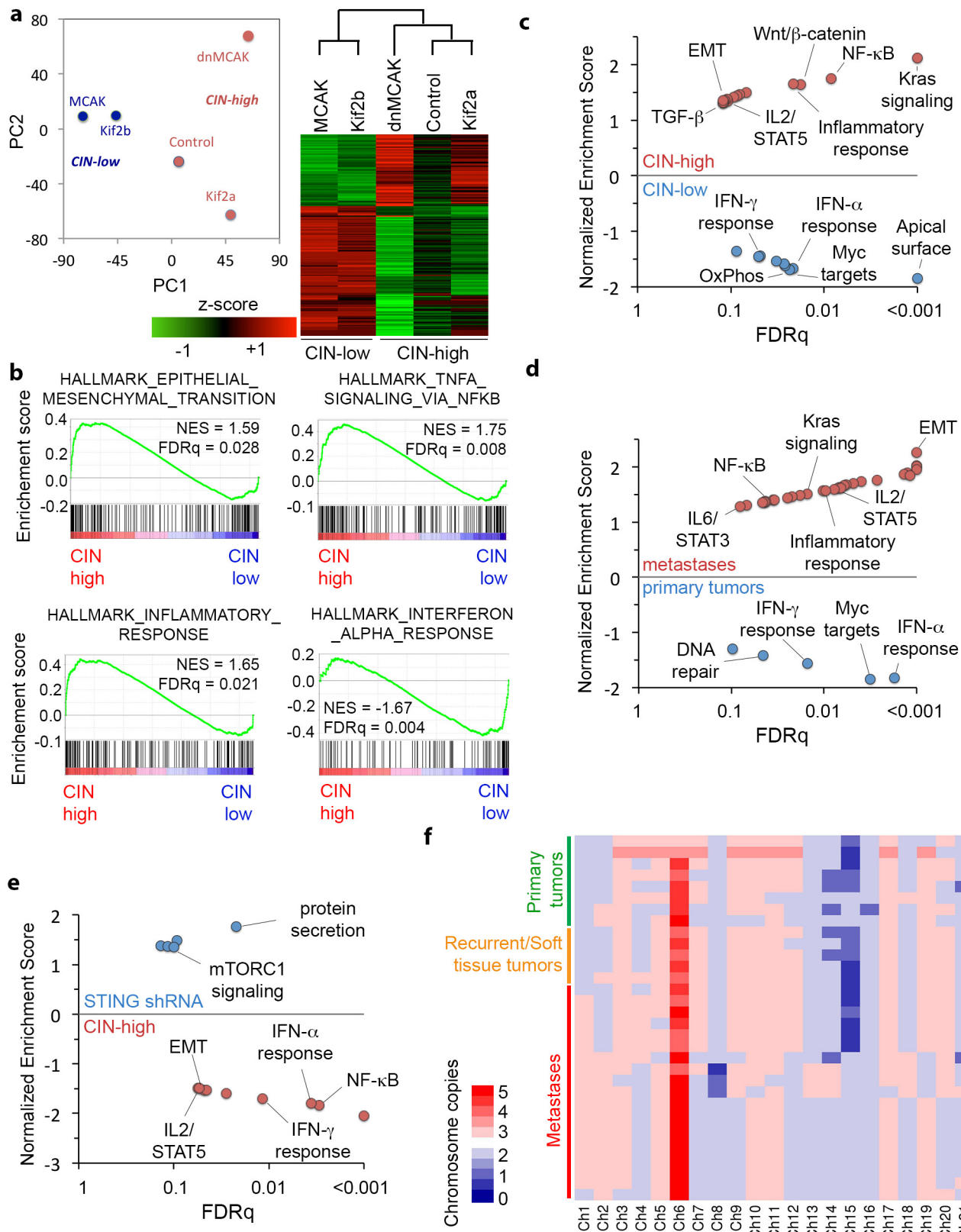


Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | CIN promotes the formation and maintenance of metastasis.

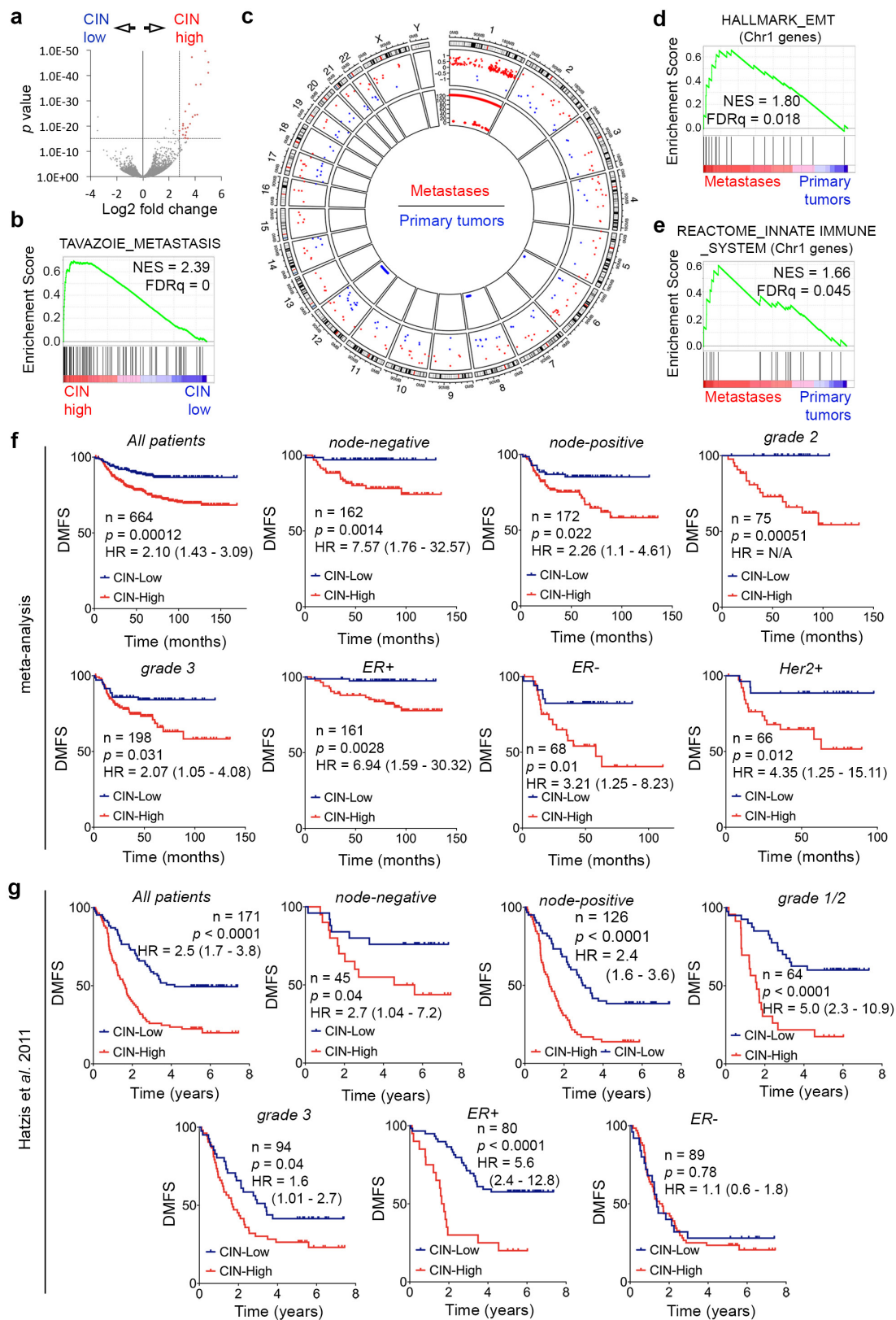
a, Normalized photon flux over time of whole animals injected with MDA-MB-231 cells expressing kinesin-13 proteins. Data shown as mean \pm s.e.m. $n = 8$ (MCAK), 7 (KIF2B), 5 (control), 4 (KIF2A), and 9 (dnMCAK) mice per group; three independent experiments performed. **b**, Representative images of mice injected with MDA-MB-231 cells expressing dnMCAK (above) or KIF2B (below) with disease burden tracked using BLI; three independent experiments performed. **c**, Photon flux (p s^{-1}) of whole animals imaged 5 weeks after intracardiac injection with control or MCAK-expressing H2030 cells. Horizontal bars represent the mean, significance tested using two-sided Mann–Whitney test, $n = 10$ mice in the MCAK group and 5 mice in the control group. **d**, Left, representative BLI images (from two independent experiments) of mice orthotopically transplanted with MDA-MB-231 cells. Images taken before (day 33) and after (day 90) tumour excision. Metastasis can be detected in the mouse transplanted with dnMCAK-expressing cells at day 90. Middle, total flux (p s^{-1}) emitted from primary tumours 52 days after transplantation. Data shown as mean \pm s.d., $n = 5$ (CIN-low) and 14 (CIN-high) mice, $P = 0.13$, two-sided Mann–Whitney test. Right, DMFS of mice orthotopically transplanted with MDA-MB-231 cells with various levels of CIN. $n = 15$ (CIN-low) and 29 (CIN-high) mice, pairwise significance tested with two-sided log-rank test. **e**, Tumour volume at 8 days (top) and survival (bottom) of mice transplanted with mouse

4T1 cells into the mammary fat pad. Bars represent median \pm interquartile range, pairwise significance tested with two-sided t -test (top) and two-sided log-rank test (bottom). $n = 20$ (CIN-low) and 30 (CIN-high) mice. **f**, Top, immunoblots of MDA-MB-231 cells overexpressing MCAK or MCAK and MAD2 stained for MAD2 using anti-MAD2 antibody with α -tubulin used as a loading control; three independent experiments performed. Bottom, percentage of anaphase cells exhibiting evidence of chromosome missegregation in cells overexpressing MCAK or MCAK and MAD2. Data shown as mean \pm s.d., $n = 150$ cells, three experiments performed, significance tested using two-sided t -test. **g**, Top, immunoblots of MDA-MB-231 cells overexpressing dnMCAK or dnMCAK and lamin B2 stained for lamin B2 using anti-lamin B2 antibody with β -actin used as a loading control. Two experiments performed. Bottom, percentage of anaphase cells exhibiting evidence of chromosome missegregation in cells overexpressing dnMCAK or dnMCAK and lamin B2. Data shown as mean \pm s.d., $n = 150$ cells, three experiments performed, significance tested using two-sided t -test. **h**, Photon flux (p s^{-1}) of whole animals after intracardiac (left) or tail vein (right) injection with MDA-MB-231 cells expressing dnMCAK or dnMCAK and lamin B2. Bars represent the median, significance tested using two-sided Mann–Whitney test, $n = 9$ (dnMCAK), 15 (dnMCAK and Lamin B2) mice in the intracardiac injection cohort and 5 mice per group in the tail vein injection cohort.



Extended Data Figure 4 | Transcriptional consequences of CIN in cancer cells. **a, b**, Principal component analysis (left) and unsupervised clustering (right) of five MDA-MB-231 cell lines expressing different kinesin-13 proteins based on bulk RNA expression data. **b–e**, Gene set enrichment analysis results showing HALLMARK gene sets that are highly enriched in CIN-high (control, KIF2A, and dnMCAK) compared with CIN-low cells (MCAK and KIF2B) (**b, c**) or STING-depleted cells (**e**),

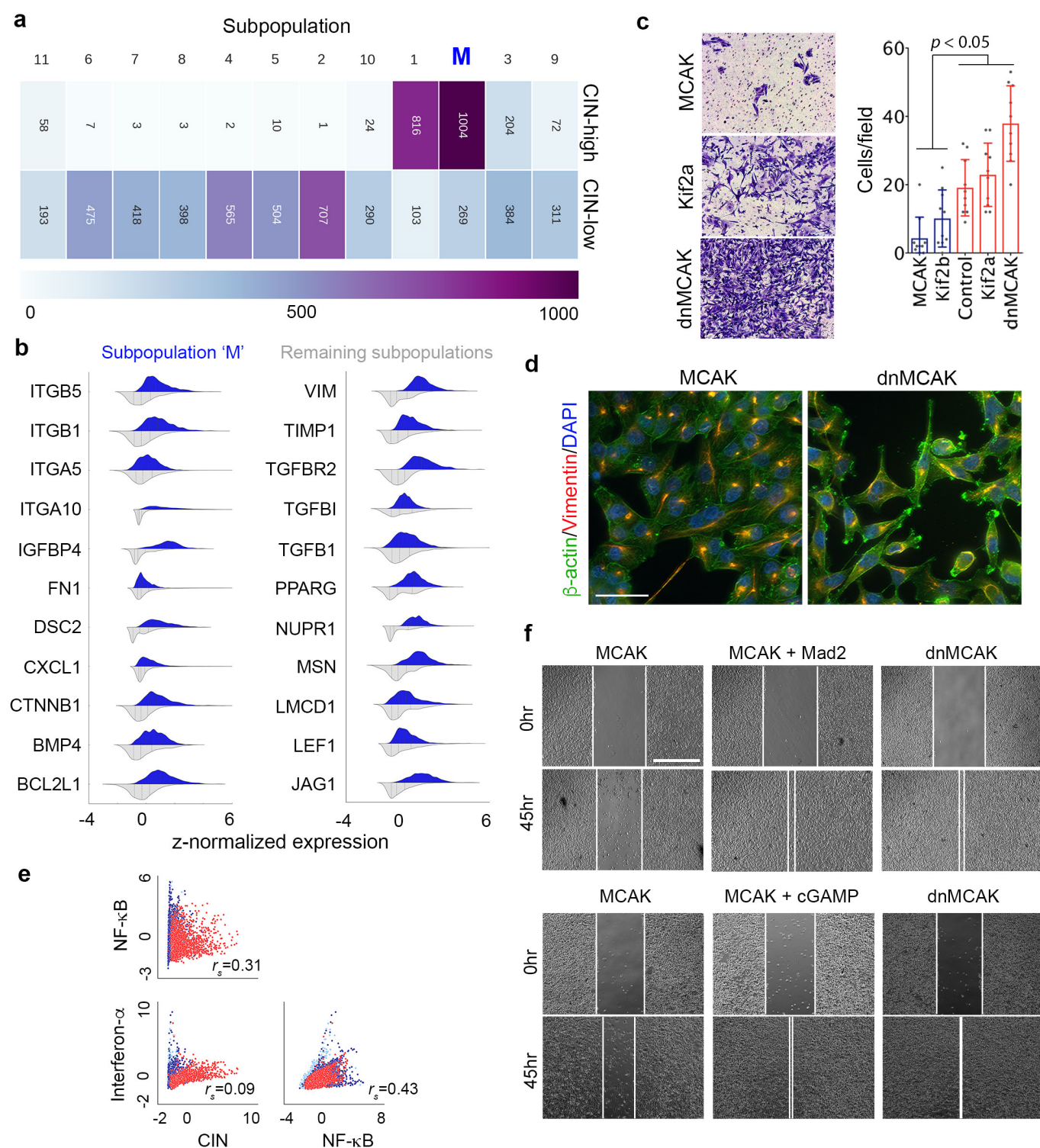
or after comparing metastases with primary tumours (**d**). Significance tested using one-sided weighted Smirnov–Kolmogorov test corrected for multiple tests. **f**, Heat map of consensus chromosomal karyotypes of cells derived from primary tumours and metastases showing selective increase in chromosome 1 copy number in metastases compared with primary tumours.



Extended Data Figure 5 | Prognostic impact of CIN signature.

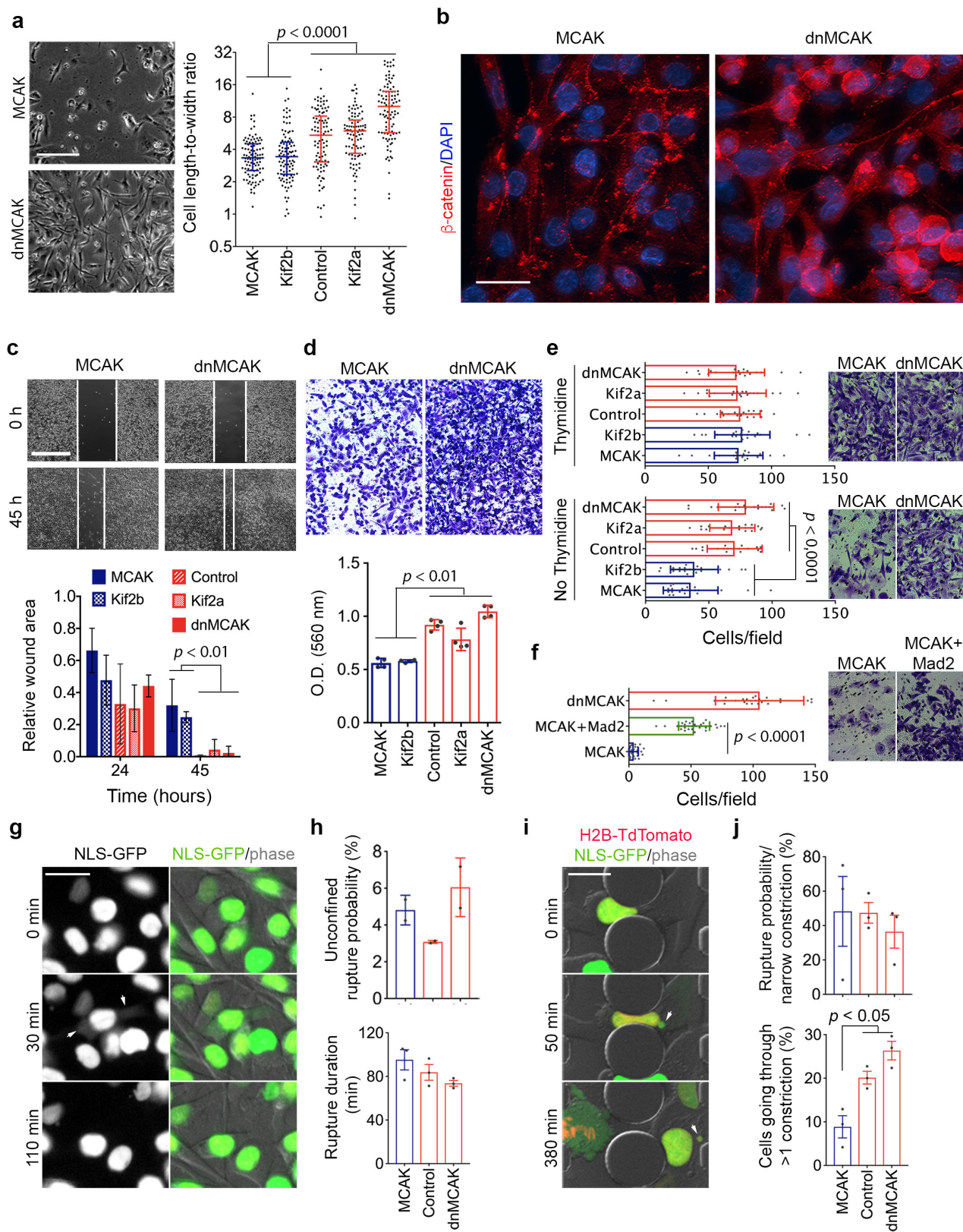
a, Volcano plot showing genes that were differentially expressed between CIN-high and CIN-low MDA-MB-231 cells. Red data points denote genes subsequently used for determining the CIN signature. **b-e**, Enrichment plots for all differentially expressed genes (**a**) or those on chromosome 1 (**d, e**). **c**, Circos plot shows genomic location (outer circle), \log_2 fold expression of genes significantly differentially expressed in metastases compared to primary tumours (middle circles), and $\log_{10} P$ (inner circle) for genomic amplifications (red) or deletions (blue) in metastases relative to primary

tumours. $n = 2$ (CIN-low), 3 (CIN-high), 11 (primary tumours), 28 (metastases). Significance tested using two-sided Wald test (**a**), one-sided weighted Smirnov-Kolmogorov test (**b, d, e**), and one-sided hypergeometric test (**c**), all corrected for multiple testing. **f, g**, DMFS of breast cancer patients stratified by lymph node status, grade, and receptor status, from a meta-analysis (**f**, $n = 664$ patients) or a validation cohort (**g**, $n = 171$ patients) divided on the basis of average expression of the CIN gene expression. Significance tested using two-sided log-rank test.



Extended Data Figure 6 | Single-cell sequencing and population detection. **a**, The cellular composition of every subpopulation presented Fig. 4b. **b**, Violin plots showing expression probability density of key metastasis and invasion genes in a subpopulation of cells ($n = 1,273$ cells) enriched for EMT and CIN genes (subpopulation M) compared with the remaining subpopulations ($n = 5,548$ cells) that were identified using graph-based unsupervised K -nearest neighbour embedding. **c**, Representative low-power field images (left) and numbers (right) of MDA-MB-231 cells that invaded through a collagen membrane within 18 h of culture. Data shown as mean \pm s.d., significance tested using

two-sided Mann–Whitney test, $n = 10$ high-power fields, two independent experiments performed. **d**, Representative images of MDA-MB-231 cells expressing MCAK or dnMCAK stained for β -actin, vimentin, and DNA. Scale bar, 50 μ m, $n = 2$ independent experiments. **e**, Single-cell correlation plots between CIN signature genes, canonical NF- κ B and type I interferon target genes, $n = 6,821$ cells. **f**, Representative phase-contrast images of a wound-healing assay of MDA-MB-231 cells expressing MCAK, MCAK and MAD2 or dnMCAK, and MCAK-expressing cells treated with cGAMP. Scale bar, 800 μ m, four experiments performed.

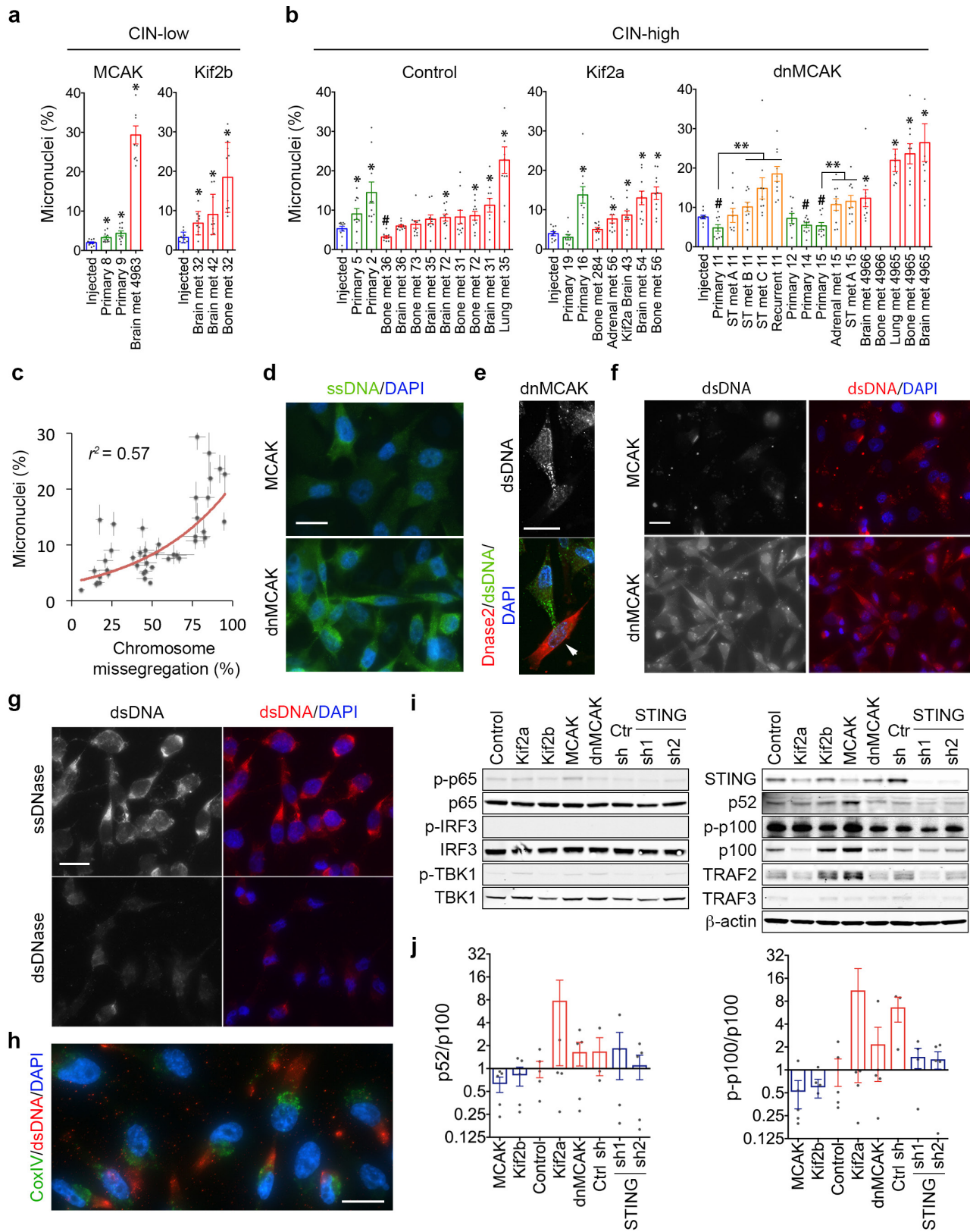


Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | CIN promotes *in vitro* invasion and

migration. **a**, Left, representative phase-contrast images of MDA-MB-231 cells in the wound area, 36 h after wound creation. Four experiments performed. Right, length-to-width ratio of cells expressing different kinesin-13 proteins. Bars span the interquartile range, $n = 100$ cells, two independent experiments performed, significance tested using two-sided Mann–Whitney test. **b**, Representative MDA-MB-231 cells stained for β -catenin (anti- β -catenin antibody) or DNA (DAPI). Changes in β -catenin are seen upon alteration of CIN; it is enriched at cell–cell junctions in MCAK-expressing cells but is found in the cytoplasm and nucleus in dnMCAK-expressing cells. Scale bar, $30\ \mu\text{m}$, two experiments performed. **c**, Top, phase-contrast images of a wound-healing assay of cells expressing kinesin-13 proteins. Scale bar, $800\ \mu\text{m}$, two experiments performed. Bottom, wound area (normalized to the 0 h time point) 24 h and 45 h after wound creation. Data shown as mean \pm s.d., $n = 4$ experiments, significance tested using two-sided t -test. **d**, Top, low-power field images of MDA-MB-231 cells that have migrated through a polycarbonate membrane containing $8\text{-}\mu\text{m}$ pores within 18 h of culture. Bottom, normalized OD of cells scraped from the bottom of the membrane. Data shown as mean \pm s.e.m., significance tested using two-sided t -test, $n = 3$ experiments. **e**, **f**, Left, number of MDA-MB-231 cells

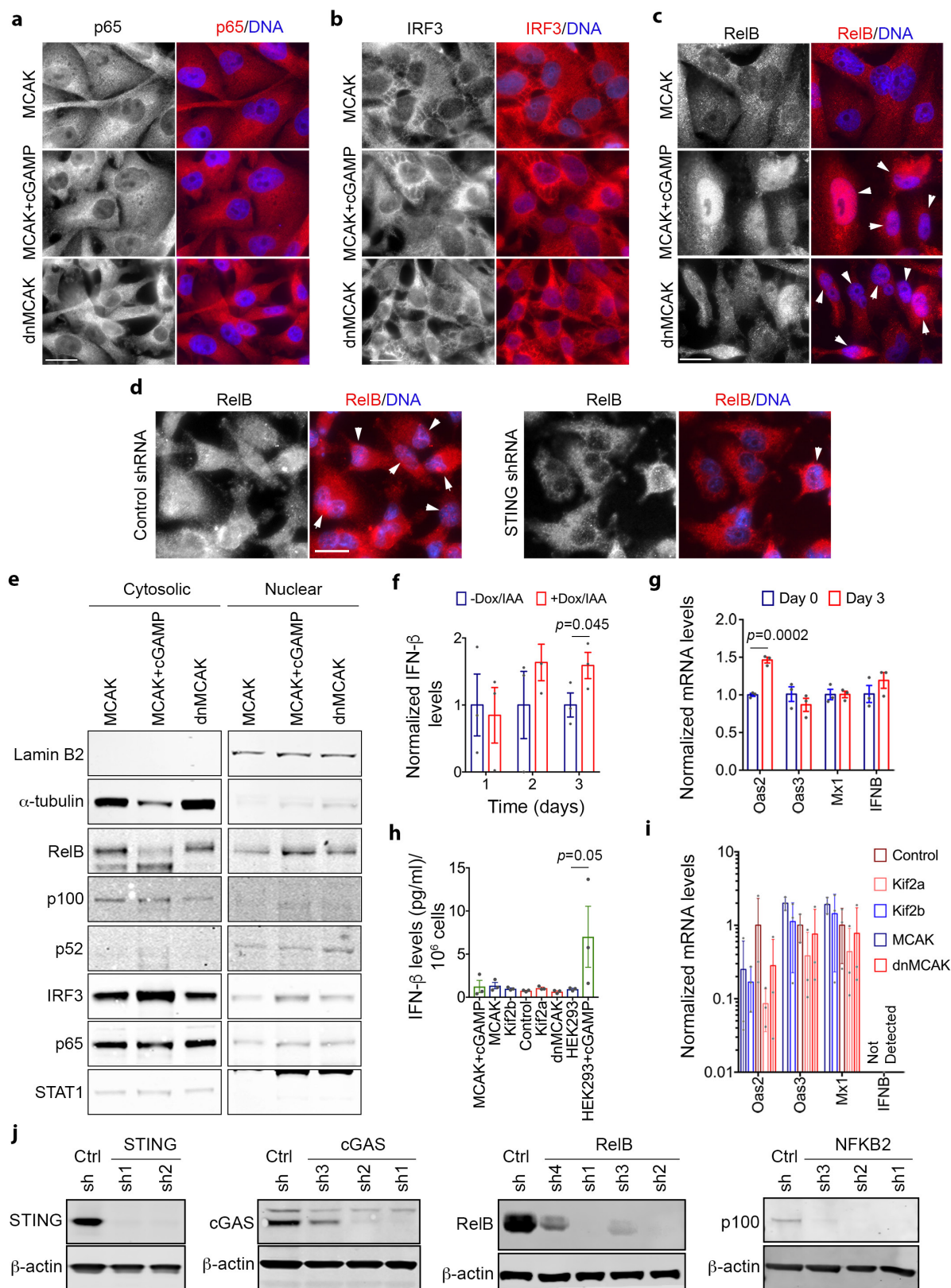
that have successfully invaded through a collagen basement membrane 24 h after plating. Data shown as mean \pm s.d., $n = 20$ high power fields from two independent experiments, significance tested using two-sided Mann–Whitney test. Right, representative images from high-power fields. Two independent experiments performed. **g**, **i**, Representative time-lapse fluorescence and phase-contrast image sequences of control cells expressing NLS–GFP undergoing unconfined migration (**g**) or going through $1 \times 5\text{-}\mu\text{m}^2$ constrictions (**i**). Scale bars, $20\ \mu\text{m}$. Arrows in **g** indicate cytoplasmic NLS–GFP. Arrows in **i** indicate formation of nuclear protrusion and subsequent fragments during confined migration. Three independent experiments performed. **h**, **j**, Top, the probability of primary nuclear rupture during unconfined conditions (**h**) or after migration through $1 \times 5\text{-}\mu\text{m}^2$ constrictions (**j**). Bottom, the number of cells migrating through more than one $1\text{-}\mu\text{m}$ -wide constrictions (**j**) and the duration of nuclear rupture (**h**), as measured by the length of time for which NLS–GFP signal is observed in the cytosol. Data shown as mean \pm s.e.m., $n = 3$ independent experiments (except for unconfined rupture probability, 2 independent experiments) encompassing 390–665 (**h**) and 150–336 (**j**) cells observed during unconfined and confined migration, respectively. Significance tested using two-sided t -test.



Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | CIN generates micronuclei and cytosolic DNA. **a, b**, Percentage of micronuclei in samples depicted in Fig. 3c, **d**: injected cells (blue), first-passage cells derived from primary tumours (green), or metastases (orange denotes spontaneous metastases arising from primary tumours, red denotes metastases obtained from direct intracardiac implantation). Data shown as mean \pm s.e.m., $n = 10$ high-power fields encompassing 500–1,500 cells per sample, three independent experiments performed, $*P < 0.05$ (denotes samples with higher missegregation rates than the injected lines), $\#P < 0.05$ (denotes samples with lower missegregation rates than the injected lines), $***P < 0.05$ (denotes significant differences between metastases and matched primary tumours from the same animals), two-tailed t -test. **c**, Correlation between the percentage of cells exhibiting evidence of chromosome missegregation and the percentage of micronuclei in all injected cell lines as well as cells derived from primary tumours and metastases. Data shown as mean \pm s.e.m., $n = 44$ samples.

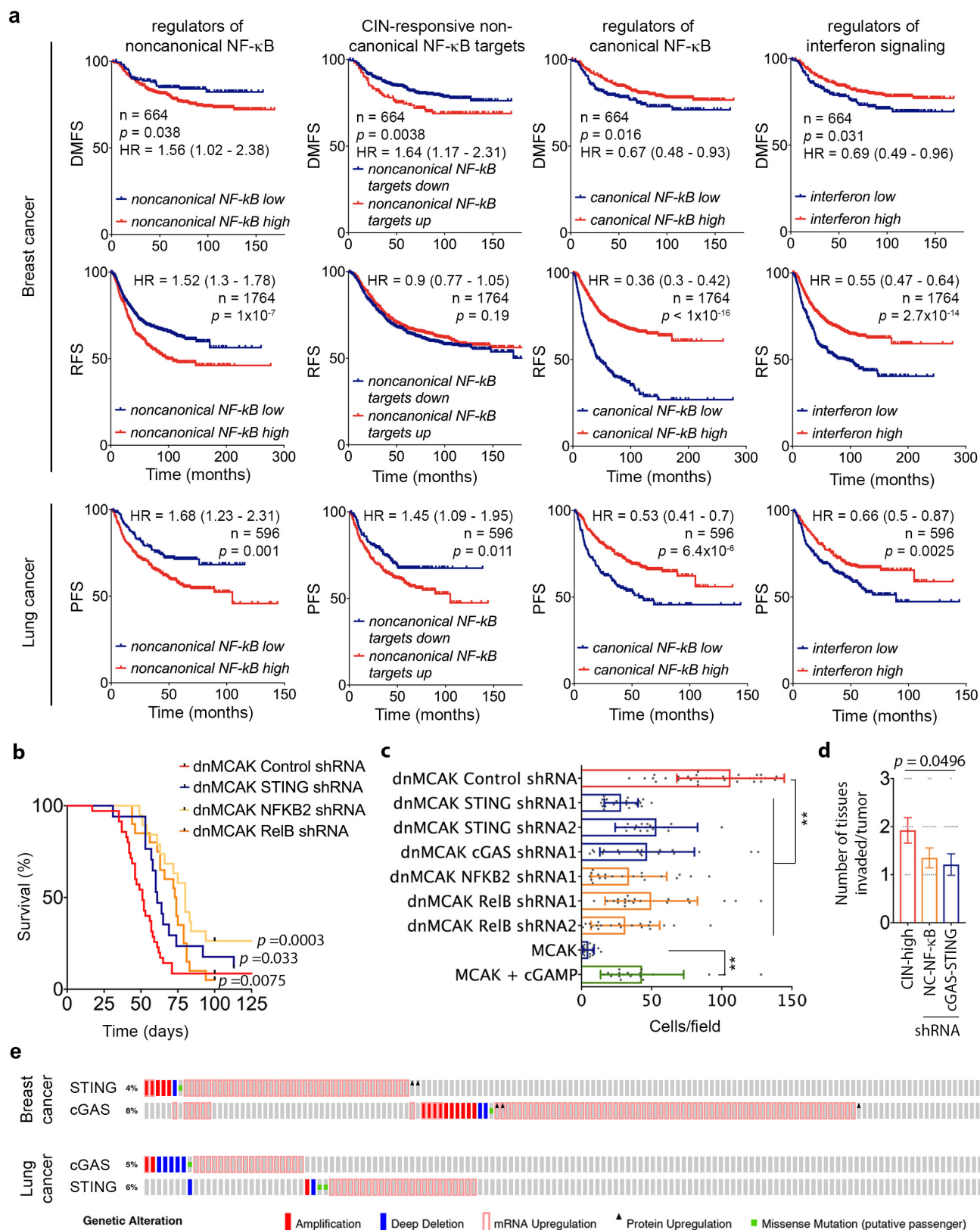
d–f, Representative images of cells stained for DNA (DAPI), cytosolic single-stranded DNA (ssDNA) (**d**), DNASE2 (RFP reporter) (**e**), or cytosolic dsDNA (**f**). Scale bar, $20\mu\text{m}$, arrows in **e** denote DNASE2-expressing cell, two independent experiments performed. **g**, Representative images of dnMCAK-expressing cells treated with ssDNASE or dsDNASE for 10 min after selective plasma membrane permeabilization (using 0.02% saponin) and stained for DNA (DAPI) and cytosolic dsDNA. Scale bar, $20\mu\text{m}$, one experiment performed. **h**, Representative images of dnMCAK-expressing cells stained for mitochondria (anti-CoxIV antibody), DNA (DAPI) or for cytosolic DNA (anti-dsDNA antibody). Scale bar, $20\mu\text{m}$, two independent experiments performed. **i**, Immunoblots of lysates from cells expressing different kinesin-13 proteins, control or STING shRNA. β -actin used as a loading control. **j**, Normalized ratio of phosphorylated p52 to p100 (left) and p100 to total p100 (right) protein levels. Data shown as mean \pm s.e.m., $n = 5$ independent experiments.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Alternative response to cytosolic DNA in cancer cells. a–d, Representative images of MDA-MB-231 cells stained for DNA (DAPI) and for p65 (**a**), IRF3 (**b**), or RELB (**c**, **d**). Images were individually contrast-enhanced to emphasize nuclear versus cytosolic localization of p65, IRF3, and RELB. For quantitative comparisons of identical images, see Supplementary Fig. 3. Arrows (**c**, **d**) point to RELB-positive nuclei. Scale bars, 20 μ m, three independent experiments performed. **e**, Immunoblots of fractionated lysates. α -tubulin and lamin B2 were used as loading controls for the cytoplasmic and nuclear fractions, respectively; three independent experiments performed. **f**, **h**, Interferon- β levels in conditioned medium from DLD-1 cells (**f**), MDA-MB-231 or

HEK293 cells with and without cGAMP addition (**h**). Data shown as mean \pm s.e.m. $n = 3$ experiments, significance tested using one-sided Mann–Whitney test. **g**, **i**, Relative levels of interferon-responsive genes obtained by RT–qPCR in DLD-1 cells (**g**) normalized to untreated conditions or MDA-MB-231 cells (**i**) normalized to control cells. Data shown as mean \pm s.d. $n = 3$ experiments, significance tested using two-sided t -test. **j**, Immunoblots of lysates of dnMCAK-expressing cells that also co-expressed control shRNA or shRNAs targeting components of the cytosolic DNA-sensing or noncanonical NF- κ B pathways. shRNA hairpins are numbered in ascending order according to the efficiency of protein knockdown. Two independent experiments performed.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | Effect of cytosolic DNA-sensing pathways on prognosis. **a**, Distant metastasis-free survival (DMFS), relapse-free survival (RFS) and progression-free survival (PFS) of patients with breast and lung, stratified according to their expression of NF- κ B and interferon pathways. Significance tested using two-sided log-rank test. **b**, Disease-specific survival of mice injected with dnMCAK-expressing MDA-MB-231 cells co-expressing control shRNA, STING shRNA, NFKB2 shRNA, or RELB shRNA. $n = 35, 16, 19$, and 20 mice in the control, STING shRNA, NFKB2 shRNA, and RELB shRNA groups, respectively; significance tested using two-sided log-rank test. **c**, Number of MDA-MB-231 cells expressing shRNA targeting genes belonging to the DNA-sensing or noncanonical

NF- κ B pathways that invaded through a collagen membrane within 24 h of culture. Data shown as mean \pm s.d., $** P < 0.0001$, two-sided Mann–Whitney test, $n = 20$ high-power fields, two independent experiments performed. **d**, Number of different normal tissues (vascular, neuronal, or soft tissue) invaded by orthotopically transplanted tumours. Data shown as mean \pm s.e.m., $*P < 0.05$, two-tailed t -test, $n = 13$ tumours (CIN-high), 20 tumours (noncanonical NF- κ B depleted), 19 tumours (cGAS-STING depleted). **e**, Oncoprints showing genomic alterations in STING (TMEM173) and cGAS (MB21D1) in breast and lung cancers from the TCGA database.

An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5

Eduardo Bañados¹, Bram P. Venemans², Chiara Mazzucchelli², Emanuele P. Farina², Fabian Walter², Feige Wang^{2,3}, Roberto Decarli^{2,5}, Daniel Stern⁶, Xiaohui Fan⁷, Fred Davies⁸, Joseph F. Hennawi⁸, Rob Simcoe⁹, Monica L. Turner^{9,10}, Hans-Walter Rix², Jinyi Yang^{3,4}, Daniel D. Kelson¹, Gwen Rudie¹ & Jan Martin Winters¹¹

Quasars are the most luminous non-transient objects known, and as such they enable studies of the Universe at the earliest cosmic epochs. Despite extensive efforts, however, the quasar ULAS J1120+0641 at redshift $z=7.09$ (hereafter J1120+0641) has remained the only one known at $z > 7$ for more than half a decade¹. Here we report observations of the quasar ULAS J134208.10+092838.61 (hereafter J1342+0928) at a redshift of $z=7.54$. This quasar has a bolometric luminosity of $4 \times 10^{14} L_{\odot}$ and a black hole mass of $8 \times 10^8 M_{\odot}$. The existence of this supermassive black hole when the Universe was only 690 million years old, just five per cent of its current age, reinforces early models of black hole growth that allow black holes with initial masses of more than about $10^4 M_{\odot}$ (refs 2, 3) or episodic hyper-Eddington accretion^{4,5}. We see strong evidence of the quasar's Ly α emission line being absorbed by a Gunn–Peterson damping wing from the intergalactic medium, as would be expected if the intergalactic hydrogen surrounding J1342+0928 is significantly neutral. We derive a significant neutral fraction, although the exact value depends on the modelling. However, even in our most conservative analysis we find $x_{\text{HI}} > 0.33$ ($x_{\text{HI}} > 0.11$) at 68 per cent (95 per cent) probability, indicating that we are probing well within the reionization epoch.

We selected the quasar J1342+0928 as part of an on-going effort to find $z > 7$ quasars by mining three large area surveys: the *Wide-field Infrared Survey Explorer*⁶ (ALLWISE), the United Kingdom Infrared Telescope Infrared Deep Sky Survey (UKIDSS) Large Area Survey⁷, and the DECam Legacy Survey (DECaLS; <http://legacysurvey.org/decamls>). At redshift $z \gtrsim 7$, residual neutral hydrogen in the intergalactic medium (IGM) absorbs virtually all flux blueward of the Ly α emission line, redshifted to observed wavelengths of $\gtrsim 1 \mu\text{m}$, making quasars drop out of the optical bands entirely. We therefore required a detection in both UKIDSS *J* and *WISE W1* bands with a signal-to-noise ratio (S/N) greater than 5 and no source in the DECaLS DR3 catalog within $3''$. We then performed forced photometry in the DECaLS z_{DE} -band image to confirm the non-detection, requiring a drop in flux of $z_{\text{DE},3\sigma} - J > 2$. We also required a flat spectral energy distribution to remove a large fraction of the most common contaminants of $z \gtrsim 7$ quasar searches⁸: low-mass brown dwarfs in our Galaxy. The survey photometry used to identify J1342+0928 is listed in Extended Data Table 1.

We confirmed J1342+0928 as a quasar with a 10 min spectrum with the Folded-port InfraRed Echellette (FIRE) spectrograph in the prism mode at the Magellan 6.5m Baade telescope at Las Campanas Observatory on 09 March 2017. To analyze the emission line properties in greater detail, we obtained deeper and higher resolution spectra with FIRE, the LBT Utility Camera in the Infrared (LUCI)

spectrograph at the Large Binocular Telescope, and the Gemini Near-Infrared Spectrograph (GNIRS) at the Gemini North telescope. The LUCI spectrum provided the first detection of the Mg II emission line at $2.4 \mu\text{m}$ but it was superseded by the higher S/N and larger wavelength coverage of the GNIRS spectrum. We also obtained deep follow-up photometry with the Magellan/Fourstar infrared camera on 19 March 2017. These data were used to bring the spectra to an absolute flux scale, compensating for slit-losses. The combined spectrum and follow-up photometry of J1342+0928 are shown in Figure 1.

The systemic redshift of this quasar is $z = 7.5413 \pm 0.0007$, measured through IRAM NOEMA observations of the [C II] $158 \mu\text{m}$ emission line from its host galaxy⁹. The redshift measured from a Gaussian fit to the Mg II line (see Figure 1) is $z = 7.527 \pm 0.004$, i.e., blueshifted by $500 \pm 140 \text{ km s}^{-1}$ with respect to the systemic redshift. This is consistent with the velocity offsets observed in other $z > 6$ quasars¹⁰. Adopting a cosmology¹¹ with $H_0 = 67.7 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.307$, and $\Omega_{\Lambda} = 0.693$, this quasar is situated at a cosmic age of just 690 Myr after the Big Bang, i.e., when the universe was $\sim 10\%$ younger than at the redshift of the previous most distant quasar known¹, at times when conditions in the universe were changing rapidly¹².

The mass of the quasar's central black hole can be estimated through the quasar luminosity and the full-width at half maximum (FWHM) of its Mg II line, under the assumption that local scaling relations¹³ are still valid at high luminosity and high redshift¹⁴. The apparent ultraviolet magnitude measured at rest-frame 1450 \AA , from the quasar spectrum is $m_{1450} = 20.34 \pm 0.04$, which translates to an absolute magnitude of $M_{1450} = -26.76 \pm 0.04$. To calculate the quasar bolometric luminosity (L_{Bol}), we first fit a power-law continuum to the spectrum and measure the luminosity at rest-frame 3000 \AA (L_{3000}). We then use the bolometric correction¹⁵ $L_{\text{Bol}} = 5.15 \times L_{3000}$, resulting in $L_{\text{Bol}} = 4 \times 10^{13} L_{\odot}$. The Mg II line has a FWHM of $2500^{+480}_{-320} \text{ km s}^{-1}$, which together with the luminosity yields a black hole mass of $7.8^{+3.3}_{-1.9} \times 10^8 M_{\odot}$. The reported errors do not include the dominant systematic uncertainties in the local scaling relations¹³ of 0.55 dex. The accretion rate of this quasar is consistent with Eddington accretion, with an Eddington ratio of $L_{\text{Bol}}/L_{\text{Edd}} = 1.5^{0.5}_{-0.4}$.

The existence of supermassive black holes in the early universe poses crucial questions on their formation and growth processes. Observationally, the most distant quasars provide joint constraints on black hole mass seed and accretion efficiency. Assuming a typical matter-energy conversion efficiency⁴ of 10%, a black hole accreting at the Eddington rate grows exponentially on timescales of $\sim 50 \text{ Myr}$. Figure 2 shows the black hole growth of three quasars assuming that they accrete at the Eddington limit during their entire life. These are the quasars that currently place the strongest constraints on early black hole growth: J1342+0928 at $z = 7.54$, J1120+0641 at $z = 7.09^1$, and

¹The Observatories of the Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, California 91101, USA. ²Max Planck Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany. ³Department of Astronomy, School of Physics, Peking University, Beijing 100871, China. ⁴Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China.

⁵INAF—Osservatorio Astronomico di Bologna, via Gobetti 93/3, 40129 Bologna, Italy. ⁶Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA. ⁷Steward Observatory, The University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721-0065, USA. ⁸Department of Physics, Broida Hall, University of California, Santa Barbara, California 93106-9530, USA. ⁹MIT-Kavli Center for Astrophysics and Space Research, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. ¹⁰Las Cumbres Observatory, 6740 Cortona Drive, Goleta, California 93117, USA. ¹¹Institut de Radioastronomie Millimétrique (IRAM), 300 rue de la Piscine, 38406 Saint Martin d'Hères, France.

SDSS J0100+2802 at $z = 6.33^{16}$. In all three cases, black hole seeds of at least $1000 M_{\odot}$ are required by $z = 40$. The existence of these super-massive black holes at $z > 7$ is at odds with early black hole formation models that do not involve either massive ($\gtrsim 10^4 M_{\odot}$) seeds or episodes of hyper-Eddington accretion.

The epoch of reionization was the universe's last major phase transition, when it changed from being completely neutral to ionized. The presence of complete Gunn-Peterson troughs in the spectra of $z \gtrsim 6$ quasars indicates that there were only traces of neutral hydrogen ($\bar{x}_{\text{HI}} > 10^{-4}$) at $z \sim 6$. Because the Ly α transition saturates at larger neutral fraction, Gunn-Peterson troughs are only sensitive to the end phases of reionization¹⁷. Therefore, to probe the epoch when reionization occurred, we need alternative methods. During earlier stages of reionization ($\bar{x}_{\text{HI}} > 0.1$), neutral intergalactic matter should produce characteristic damped Gunn-Peterson absorption *redward* of the Ly α emission line¹⁸. Evidence of this long-sought signature in quasar spectra has only been reported once, in the previous redshift-record holder at $z = 7.09^{1,19,20}$.

To calculate the implied IGM neutral fraction \bar{x}_{HI} , one must first estimate the shape of the unabsorbed continuum, and then fit a parameterized absorption model using the data and continuum as inputs. This analysis is challenging because assumptions about the process of reionization need to be made and estimating the intrinsic strength of the Ly α emission for one single quasar is not straightforward. The latter is particularly difficult for the case of J1342+0928, which has extreme line blueshifts, which greatly reduces the number of lower-redshift quasars with which this source can be compared. In Figure 3 we summarize our approach, where we have followed previous works^{1,20} to estimate the intrinsic continuum by searching for lower redshift quasars with similar spectral features and obtained an estimate of the neutral fraction following the method outlined by Miralda-Escudé (1998)¹⁸. The main result is that a significantly neutral IGM is required to reproduce the Ly α damping wing profile of J1342+0928. We find $\bar{x}_{\text{HI}} = 0.56^{+0.21}_{-0.18}$, with the 95% central interval of the \bar{x}_{HI} distribution being between 0.26 and 0.93 (see Figure 3). To explore how robust this result is, we introduce an alternative quasar intrinsic emission model and two more elaborated methods to model the IGM damping wing in the Methods section. All our analyses strongly favour a scenario where the IGM surrounding J1342+0928 is significantly neutral although the exact value depends on the method (see Extended Data Table 2). Nevertheless, even our most conservative case indicates $\bar{x}_{\text{HI}} > 0.11$ at 95% probability. We emphasize that higher S/N and larger wavelength coverage of the quasar's spectrum would be critical to refine and strengthen this result.

An important caveat is that a similar absorption profile could also be caused by a single high column density absorber ($N_{\text{HI}} > 2 \times 10^{20} \text{ cm}^{-2}$) in the immediate vicinity of the quasar¹⁸. Although we find a large number of foreground heavy-element absorbers at lower redshifts, we find no evidence for metal-line absorption at redshifts near that of the quasar. Adopting the methodology of Simcoe et al. (2012)²¹, a single Ly α absorber at $z = 7.49 \pm 0.01$ and $\log N_{\text{HI}}/\text{cm}^{-2} = 20.53^{+0.32}_{-0.45}$ could produce the damping wing observed in J1342+0928. However, this absorber could have at most a metal abundance of 1/4500 the Solar value for oxygen (95% confidence), which would make it the most distant and metal-poor absorber known²¹. The probability of intercepting a discrete absorber with $N_{\text{HI}} > 10^{20.50} \text{ cm}^{-2}$ within 2000 km s^{-1} of a quasar at $z = 7.5$ is less than 1%, based on the number density of such systems at lower redshifts²². This low probability supports the hypothesis that the absorption profile in J1342+0928 is instead probing the neutral IGM gas in the epoch of reionization.

Finally, the fact that both quasars known at $z > 7$ present evidence of Ly α damping wings confirms that we are starting to probe well within the epoch of reionization (see Figure 4) in agreement with recent indications based on the number density of Ly α emitting galaxies at similar redshifts²³ and results from the Cosmic Microwave Background¹².

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 June; accepted 28 November 2017.

Published online 6 December 2017.

1. Mortlock, D. J. et al. A luminous quasar at a redshift of $z = 7.085$. *Nat.* **474**, 616–619 (2011).
2. Latif, M. A., Schleicher, D. R. G., Schmidt, W. & Niemeyer, J. Black hole formation in the early Universe. *Mon. Not. R. Astron. Soc.* **433**, 1607–1618 (2013).
3. Alexander, T. & Natarajan, P. Rapid growth of seed black holes in the early universe by supra- exponential accretion. *Sci.* **345**, 1330–1333 (2014).
4. Pacucci, F., Volonteri, M. & Ferrara, A. The growth efficiency of high-redshift black holes. *Mon. Not. R. Astron. Soc.* **452**, 1922–1933 (2015).
5. Inayoshi, K., Haiman, Z. & Ostriker, J. P. Hyper-Eddington accretion flows on to massive black holes. *Mon. Not. R. Astron. Soc.* **459**, 3738–3755 (2016).
6. Wright, E. L. et al. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *Astron. J.* **140**, 1868–1881 (2010).
7. Lawrence, A. et al. The UKIRT Infrared Deep Sky Survey (UKIDSS). *Mon. Not. R. Astron. Soc.* **379**, 1599–1617 (2007).
8. Bañados, E. et al. The Pan-STARRS1 Distant $z > 5.6$ Quasar Survey: More than 100 Quasars within the First Gyr of the Universe. *Astrophys. J. Suppl. Ser.* **227**, 11 (2016).
9. Venemans, B. P. et al. Copious amounts of dust and gas in a $z = 7.5$ quasar host galaxy. *Astrophys. J.* Accepted (2017).
10. Venemans, B. P. et al. Bright [CII] and Dust Emission in Three $z > 6.6$ Quasar Host Galaxies Observed by ALMA. *Astrophys. J.* **816**, 37 (2016).
11. Planck Collaboration et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.* **594**, A13 (2016).
12. Planck Collaboration et al. Planck intermediate results. XLVII. Planck constraints on reionization history. *Astron. Astrophys.* **596**, A108 (2016).
13. Vestergaard, M. & Osmer, P. S. Mass Functions of the Active Black Holes in Distant Quasars from the Large Bright Quasar Survey, the Bright Quasar Survey, and the Color-selected Sample of the SDSS Fall Equatorial Stripe. *Astrophys. J.* **699**, 800–816 (2009).
14. Mejía-Restrepo, J. E., Lira, P., Netzer, H., Trakhtenbrot, B. & Capellupo, D. M. The effect of nuclear gas distribution on the mass determination of supermassive black holes. *Nature Astronomy*. DOI: 10.1038/s41550-017-0305-z (2017).
15. Richards, G. T. et al. Spectral Energy Distributions and Multiwavelength Selection of Type 1 Quasars. *Astrophys. J. Suppl. Ser.* **166**, 470–497 (2006).
16. Wu, X.-B. et al. An ultraluminous quasar with a twelve-billion-solar-mass black hole at redshift 6.30. *Nat.* **518**, 512–515 (2015).
17. Becker, G. D. et al. Evidence of patchy hydrogen reionization from an extreme Ly α trough below redshift six. *Mon. Not. R. Astron. Soc.* **447**, 3402–3419 (2015).
18. Miralda-Escudé, J. Reionization of the Intergalactic Medium and the Damping Wing of the Gunn-Peterson Trough. *Astrophys. J.* **501**, 15–22 (1998).
19. Greig, B., Mesinger, A., Haiman, Z. & Simcoe, R. Are we witnessing the epoch of reionization at $z = 7.1$ from the spectrum of J1120+0641? *Mon. Not. R. Astron. Soc.* (2016).
20. Bosman, S. E. I. & Becker, G. D. Re-examining the case for neutral gas near the redshift 7 quasar ULAS J1120+0641. *Mon. Not. R. Astron. Soc.* **452**, 1105–1111 (2015).
21. Simcoe, R. A. et al. Extremely metal-poor gas at a redshift of 7. *Nat.* **492**, 79–82 (2012).
22. Songaila, A. & Cowie, L. L. The Evolution of Lyman Limit Absorption Systems to Redshift Six. *Astrophys. J.* **721**, 1448–1466 (2010).
23. Zheng, Z.-Y. et al. First Results from the Lyman Alpha Galaxies in the Epoch of Reionization (LAGER) Survey: Cosmological Reionization at $z \sim 7$. *Astrophys. J.* **842**, L22 (2017).
24. Gallerani, S., Fan, X., Maiolino, R. & Pacucci, F. Physical Properties of the First Quasars. *Public. Astron. Soc. Aust.* **34**, e022 (2017).
25. Pâris, I. et al. The Sloan Digital Sky Survey Quasar Catalog: Twelfth data release. *Astron. Astrophys.* **597**, A79 (2017).
26. Fan, X. et al. Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with $z \sim 6$ Quasars. II. A Sample of 19 Quasars. *Astron. J.* **132**, 117–136 (2006).
27. Eilers, A.-C. et al. Implications of $z \sim 6$ Quasar Proximity Zones for the Epoch of Reionization and Quasar Lifetimes. *Astrophys. J.* **840**, 24 (2017).

Acknowledgements We thank D. Ossip for his support with the FIRE echelle observations and A. Stephens for his help preparing the GNIRS observations. This work is based on data collected with the Magellan Baade telescope, the Gemini North telescope (program GN-2017A-DD-4), the Large Binocular Telescope, and the IRAM/NOEMA interferometer. We are grateful for the support provided by the staff of these observatories. We acknowledge the use of the UKIDSS, WISE, and DECaLS surveys.

Author Contributions E.B., R.D., X.F., E.P.F., C.M., H.-W.R., D.S., B.P.V., F.W., F.W., and J.Y. discussed and planned the candidate selection, observing strategy, and analyzed the data. E.B. selected the quasar and with D.S. took and analyzed the discovery spectrum. R.S. provided the final FIRE data reduction. J.F.H.

provided the final GNIRS data reduction. G.R. carried out the follow-up Fourstar observations for this quasar. D.K. reduced the follow-up Fourstar data. J.M.W., B.P.V., and F.W. contributed with the observations and analysis of the IRAM/NOEMA data. The Ly α damping wing analyses were carried out by E.B. (main manuscript, Model A), F.D. and J.F.H. (Methods section, Model B) and R.S. and M.L.T. (Methods section, Model C). F.D. and J.F.H. did the PCA continuum modeling described in the Methods section. R.S. performed the analysis to find the characteristics of a single absorber that could cause the quasar's absorption profile. E.B. led the team and prepared the manuscript. All co-authors discussed the results and provided input to the paper and data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to E.B. (ebanados@carnegiescience.edu).

Reviewer Information *Nature* thanks D. Mortlock and the other anonymous reviewer(s) for their contribution to the peer review of this work.

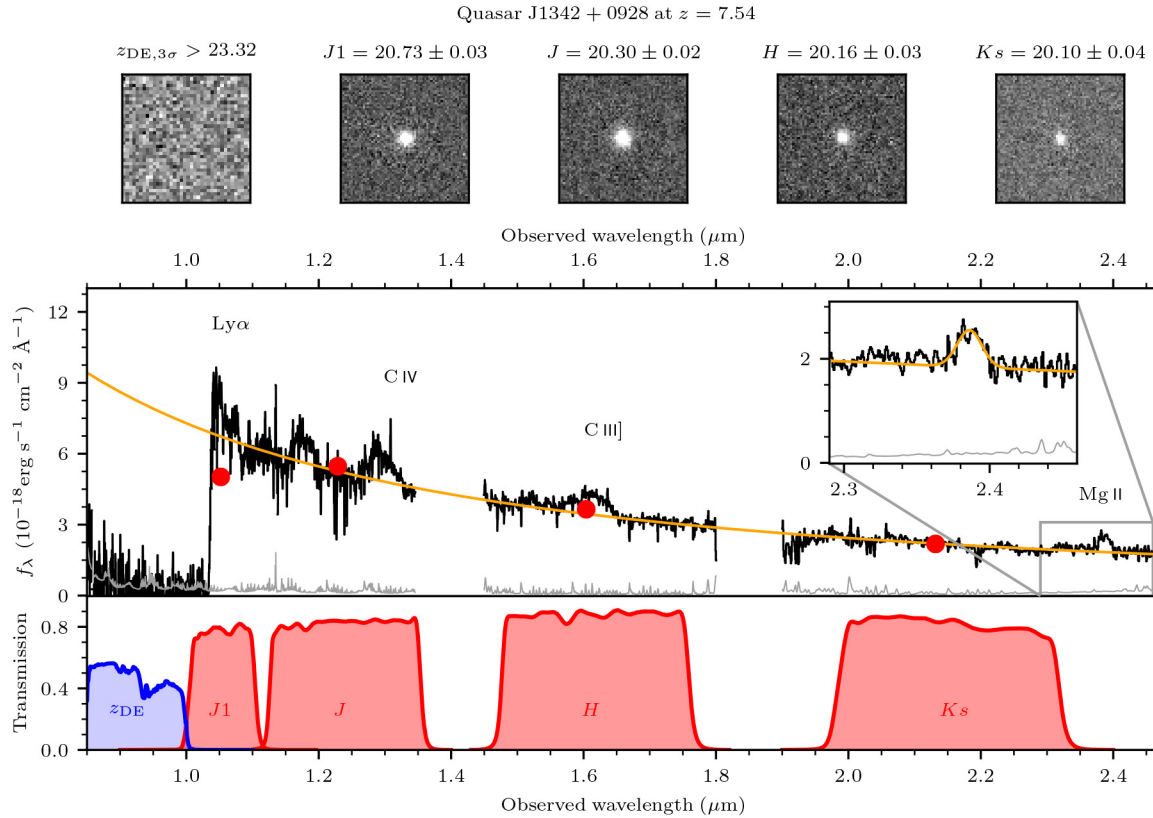


Figure 1 | Photometry and combined Magellan/FIRE and Gemini/GNIRS near-infrared spectrum of the quasar J1342+0928 at $z = 7.54$.

The FIRE data were taken on 11–12 March 2017 for a total integration time of 3.5 hr. We used the $0.6''$ slit in the echellete mode, yielding a spectral resolution of $R \sim 6000$ over the range 0.8 – $2.3 \mu\text{m}$. The GNIRS spectrum was obtained on 31 March 2017 and 03 April 2017 with a total exposure time of 4.7 hr. We used the $0.675''$ slit in the cross-dispersion mode, yielding a spectral resolution of $R \sim 1800$ over the range 0.8 – $2.5 \mu\text{m}$. The spectra are shown at the GNIRS resolution binned by a factor of two.

The 1σ error is shown in gray and the orange line represents the best-fit power-law continuum emission with $f_{\lambda} \propto \lambda^{-1.58 \pm 0.02}$. Regions with low sky transparency between the J/H and H/Ks bands are not shown. The red circles show the follow-up photometry taken with the Magellan/Fourstar infrared camera. The inset shows a Gaussian fit to the Mg II line, from which we derive a black hole mass of $7.8 \times 10^8 M_{\odot}$. The bottom panel shows the transmission of the Fourstar $J1$, J , H , Ks (red), and the DECam z_{DE} (blue) filters, while the top panel shows $10'' \times 10''$ postage stamps of the quasar in the same filters with their respective AB magnitudes.

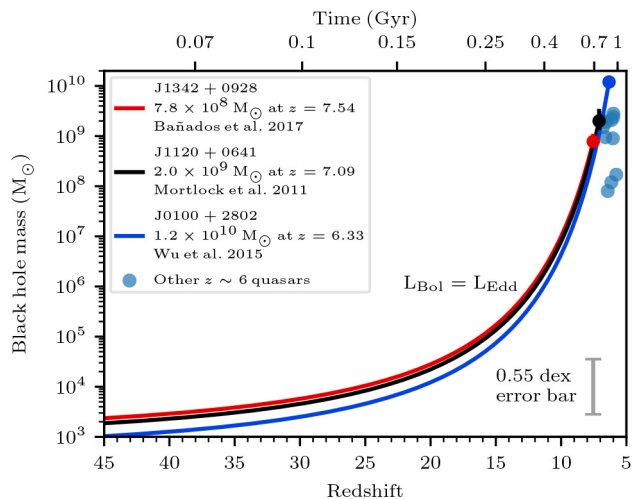


Figure 2 | Black hole growth of three of the highest redshift and most massive quasars in the early universe, J1342+0928, J1120+0641¹, and J0100+2802¹⁶. The three curves are normalized to the observed black hole mass and redshift of these quasars (data points with statistical error bars). The black hole growth is modeled as $M_{BH} = M_{BH,seed} \times \exp(\text{time}/50 \text{ Myr})$, where we have assumed that the black holes are accreting at the Eddington limit ($L_{\text{Bol}} = L_{\text{Edd}}$) with a radiative efficiency of 10%. The circles show a compilation of black hole masses of $z \sim 6$ quasars²⁴. The gray error bar at the bottom right represents dominant uncertainty due to systematics in the local scaling relation used to estimate the black hole mass of quasars at these redshifts¹³. Ignoring this systematic uncertainty and assuming that the local relations apply to these extremely distant and luminous quasars, black hole mass seeds more massive than $1000 M_{\odot}$ by $z = 40$ are required to grow the observed supermassive black holes in all three cases.

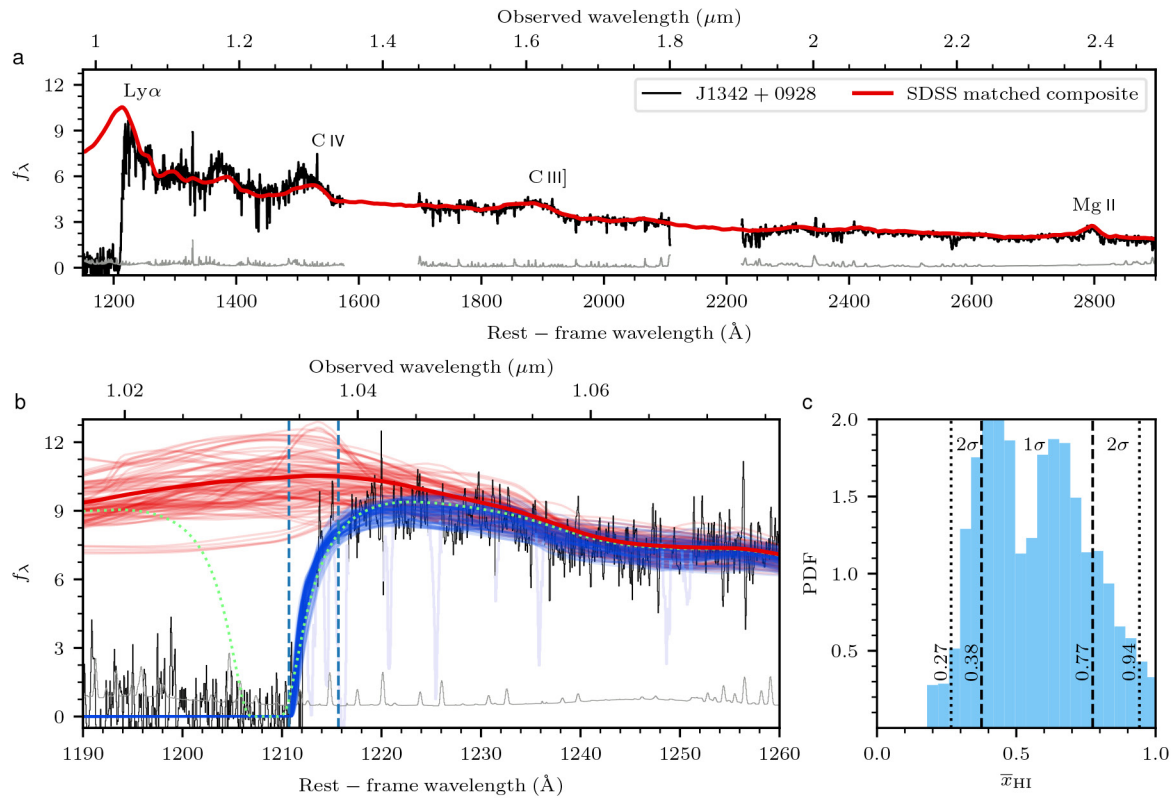


Figure 3 | Continuum emission and Ly α damping wing modeling in the spectrum of J1342+0928 (in units of $10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Å}^{-1}$). **a**, The quasar spectrum is the same shown in Figure 1. The continuum model in red is constructed by averaging the SDSS DR12 quasars²⁵ not flagged as broad-absorption line quasars in the redshift interval $2.1 < z_{\text{Mg II}} < 2.4$, and with a median $S/N > 5$ in the C IV region. The subsample was further refined considering only SDSS quasars with a C IV blueshift with respect to Mg II within 1000 km s^{-1} of that observed in J1342+0928 ($6090 \pm 275 \text{ km s}^{-1}$) and with rest-frame C IV equivalent widths consistent at the 3σ -level with the C IV equivalent width of J1342+0928 ($\text{EW} = 11.3 \pm 0.8 \text{ Å}$). This yielded 46 ‘analog’ quasars. Their continua were individually fit by a slow-varying spline to remove strong absorption systems, noisy regions, and interpolate between high transmission peaks blueward of Ly α . We then normalized each spectrum at 1290 Å and averaged them. This composite spectrum is shown in red, and reproduces fairly well the spectral features of J1342+0928. Assuming the systemic redshift derived from [C II], the proximity zone of J1342+0928, which is defined^{26,27} as the physical radius at which the transmission drops to 10%, is 1.3 Mpc . **b**, Zoom-in to the Ly α region showing a strong absorption profile that can be modeled as a Ly α damping wing caused by a significantly neutral IGM. The black and gray curves are the FIRE spectrum and 1σ error binned by a factor of two. The

pale gray lines show regions masked out due to intervening absorption systems. The thick red line is the SDSS matched composite spectrum (see panel a) while the thin red lines are used as intrinsic emission models (100 out of 3000 shown). To take into account the error in the prediction for any given quasar, these models consist of bootstrapped mean composite spectra plus a relative error vector (SDSS quasar - mean)/mean randomly chosen among the 46 possible error vectors. The blue lines represent the expected IGM damping wing following the prescription of Miralda-Escudé (1998)¹⁸ assuming a fully ionized proximity zone (region between the vertical dashed lines), a constant neutral fraction \bar{x}_{HI} between the end of the quasar’s proximity zone and $z = 7.0$, and a fully ionized IGM at $z < 7$. The exact choice of the transition redshift does not significantly affect the results. The green dotted line is the absorption that would be caused by a single absorber with $N_{\text{HI}} = 10^{20.53} \text{ cm}^{-2}$ at $z = 7.49$. **c**, The derived \bar{x}_{HI} probability density function (PDF), with a preferred value of $\bar{x}_{\text{HI}} = 0.56^{+0.21}_{-0.18}$. In the Methods section we present one additional model of the quasar’s intrinsic emission and two additional models of the IGM damping wing. All our analyses require a significantly neutral IGM to reproduce the damping wing profile in the spectrum of J1342+0928 (see Extended Data Table 2).

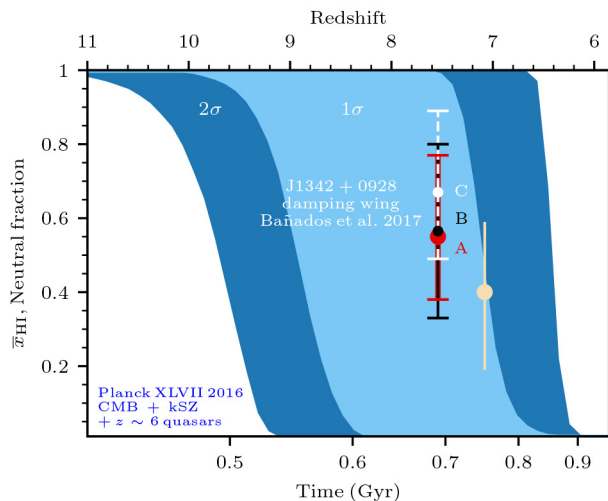


Figure 4 | Constraints on the history of reionization in terms of hydrogen neutral fraction vs. cosmic time from the Big Bang in gigayears (redshift on the top axis). The contours show the 1σ and 2σ constraints from the optical depth to the Cosmic Microwave Background (CMB), the kinetic Sunyaev-Zeldovich (kSZ) effect, and $z \sim 6$ quasars compiled by the Planck collaboration¹² (see their Figure 17 and references therein). The data points represent the 1σ constraints from the damping wing analyses of J1120+0641 at $z=7.09^{19}$ ($\bar{x}_{\text{HI}} = 0.40^{+0.21}_{-0.19}$) and J1342+0928 at $z=7.54$. The red error bar represents the analysis presented in Figure 3 (Model A), while the black and white measurements are the most conservative constraints (i.e., lower neutral fraction) from the two additional models (B and C) of the Ly α damping wing of J1342+0928 presented in the Methods section (see Extended Data Table 2). All our analyses consistently find a large fraction of neutral hydrogen surrounding J1342+0928. The uncertainties in the IGM damping wing analyses do not include cosmic variance, i.e., we are constraining one line-of-sight per quasar. In order to better comprehend the global history of reionization would require additional measurements at similar redshifts along more lines of sight. We note that the quasar IGM damping wing constraints shown in this figure have used a range of methods to perform the analysis and thus have different systematics.

METHODS

Ly α damping wing modeling. As discussed in the main manuscript, in order to calculate the intergalactic medium (IGM) neutral fraction (\bar{x}_{HI}), one must first estimate the quasar's intrinsic emission. After that, a model is fit that reproduces the observed damping wing in the Ly α region. In the main manuscript, we show that the data strongly suggests that the IGM in the surrounding of the quasar J1342+0928 at $z=7.54$ is significantly neutral: $\bar{x}_{\text{HI}} = 0.56^{+0.21}_{-0.18}$ (see Figure 3). Recovering the intrinsic continuum of J1342+0928 is particularly challenging because of its extreme emission line blueshifts, which are not well represented in matched composites at lower redshifts. Specifically, in the analysis presented in the manuscript we found only 46 lower-redshift quasars with similar C IV blueshifts and equivalent widths. Furthermore, in the main manuscript we have modeled the IGM Gunn-Peterson damping wing following a prescription¹⁷ that assumes a homogeneous neutral IGM between the quasar's proximity zone and an arbitrary redshift, which we set to $z_{\text{END}} = 7.0$ and considers the universe to be completely ionized below that redshift. We remark that given that the damping wing is produced by the IGM in the environment of the quasar, the results are insensitive to the exact value of z_{END} . The results change by less than 1% even in the extreme case of $z_{\text{END}} = 6.0$. Hereafter, we refer to the IGM model presented in the main manuscript as Model A.

Here we reanalyze the data presenting an additional model of the quasar's intrinsic emission using a principal component analysis (PCA) decomposition. Furthermore, we introduce two completely independent methods to model the IGM damping wing using the quasar's intrinsic emission obtained in the main manuscript (SDSS matched composite, hereafter continuum 1) and the PCA-model (hereafter continuum 2; described in more detail below) as inputs. In this way, we test how sensitive our conclusions are to different assumptions and systematics in the analysis.

Quasar's intrinsic emission - PCA model. In this model, we predict the intrinsic quasar continuum in the Ly α region ($\lambda_{\text{rest}} < 1260 \text{ \AA}$) based on the rest of the spectrum using a PCA analysis trained on 12764 quasar spectra from the SDSS/BOSS DR12 quasar catalog²⁵. Quasars in the training set were chosen to have $S/N > 7$ at $\lambda_{\text{rest}} < 1285 \text{ \AA}$, and to lie at redshifts $2.1 < z < 2.5$ such that the BOSS spectral range comfortably covered the Ly α region (down to $\lambda_{\text{rest}} = 1190 \text{ \AA}$) and the Mg II region up to $\lambda_{\text{rest}} \sim 2800 \text{ \AA}$, similar to the rest-frame spectral coverage of J1342+0928. We constructed red-side and blue-side PCA basis spectra from the training set after applying automated spline fits to recover the unabsorbed continuum. For each quasar in the training set, we simultaneously fit for the PCA coefficients and a template redshift, placing each quasar into a common "PCA redshift frame" independent of (but largely similar to) their catalog redshifts. We used these fits to define the matrix projection from red-side to blue-side PCA coefficients as in Suzuki et al. (2005)²⁸ and P  ris et al. (2011)²⁹. We then fit the red-side PCA coefficients of J1342+0928 in the same way as the training set. To estimate the bias and uncertainty of the continuum fit for quasars similar to J1342+0928, we measured the relative continuum error for the 1% of quasars in the training set with the most similar red-side PCA coefficients. The relative uncertainty in the Ly α region for this subset was found to be $\sim 7\%$.

In Figure E1 we compare the quasar's intrinsic Ly α emission reconstructed by both the PCA and SDSS-matching analyses. The PCA continuum predicts a slightly stronger emission in the Ly α region but in both cases the emission is significantly weaker than for an average low-redshift SDSS quasar.

IGM damping wing. Using continuum 2 (PCA) as input for Model A, we require a virtually completely neutral universe, $\bar{x}_{\text{HI}} \sim 1$, to model the Ly α damping wing in the spectrum of J1342+0928. This is driven by the higher intrinsic flux in the Ly α region of continuum 2. We would obtain an even more dramatic result if we naively used an average SDSS quasar as input, which would have a much stronger Ly α emission line (Figure E1). Altogether, this seems to reinforce our findings that the universe is significantly neutral at $z \sim 7.5$. To assess whether this result is dependent on the method used to reproduce the IGM damping wing, we here introduce two more elaborated methods (hereafter Models B and C).

Model B. In this model, we place the quasar within simulated massive dark matter halos at $z=7.5$, which in turn populate large-scale overdensities already ionized by galaxies before the quasar turns on³⁰. The damping wing strength is then sensitive to the distance to the nearest neutral patch set by the morphology of reionization, as well as the output of ionizing photons from J1342+0928. Small proximity zones can result either from a high \bar{x}_{HI} in the surrounding IGM, or from a short active lifetime of the central quasar²⁷. We modeled the residual H I absorption inside the quasar proximity zone and the damping wing profile from the neutral IGM by simulating the radiative transfer of ionizing photons through the IGM³¹ from the locations of massive dark matter halos using realistic distribution of densities from a large-volume hydrodynamical simulation³². The inside-out morphology of reionization as a function of \bar{x}_{HI} was computed from independent large-volume semi-numerical simulations of patchy reionization using a modified version of the 21cmFAST code³³.

The hydrogen neutral fraction \bar{x}_{HI} and quasar lifetime t_Q were then jointly constrained via a Bayesian approach using pseudo-likelihood in the spirit of indirect inference methods (e.g., Drovandi et al. 2015³⁴). We define our pseudo-likelihood as the product of independent flux PDFs evaluated in 500 km s^{-1} spectral bins, and treat the set of maximum pseudo-likelihood parameter values (i.e., \bar{x}_{HI} and t_Q) as a summary statistic.

We applied Model B to the quasar continuum obtained from the PCA method described above, employing millions of forward-modeled mock spectra, including a self-consistent treatment of the highly covariant continuum (determined from our PCA training set described above), to compute the posterior PDF. Marginalizing over quasar lifetimes with a log-uniform prior in the range $10^5 < t_Q < 10^8$ years, the central 68% (95%) credible interval for \bar{x}_{HI} is $0.45 - 0.87$ ($0.22 - 0.98$). We show a representation of Model B using the PCA-model intrinsic emission and its marginalized posterior PDF of \bar{x}_{HI} in Figure E2.

If we apply Model B using the SDSS matched intrinsic emission models as input, the central 68% (95%) credible interval for \bar{x}_{HI} is $0.33 - 0.80$ ($0.11 - 0.96$) and we show the posterior PDF of \bar{x}_{HI} in Figure E3a.

Model C. In this model, the IGM absorption profile is modeled based on the methods outlined in Bolton et al. (2011)³⁵. After the quasar turns on, it evacuates an expanding ionized region (the proximity zone) within the surrounding IGM. Its absorption profile is then specified by four parameters: (1) the quasar's ionizing luminosity (constrained by photometry), (2) the proximity zone size (related to the quasar age), (3) the mean density of the surrounding medium, and (4) \bar{x}_{HI} outside the proximity zone. We use an affine-invariant Markov Chain Monte Carlo (MCMC) solver³⁶ to fit these four parameters. The quasar's ionizing luminosity is estimated using the power law indices from Telfer et al. (2002)³⁷ and the quasar magnitudes from this work. We impose a Gaussian prior with a width determined from the power law index errors from Telfer et al. (2002)³⁷, while the remaining parameters are all given flat priors.

Before fitting our model, we apply an automated clipping procedure to remove spectral absorption features. We divide the normalized spectrum into bins of size 2.5 \AA in the rest-frame, and interpolate a B-spline through the mean flux in each bin. Any pixels with flux values $> 3\sigma$ below or $> 7\sigma$ above the interpolated values are masked. This procedure is repeated until convergence is achieved. Then, we ran two MCMC realizations, one for each mean quasar continuum created above (1 and 2, see Figure E1). For each continuum model, we ran 100 chains of 2000 steps each, and use the final 200 steps to construct the posteriors (the burn-in occurs within the first ~ 250 steps).

For continuum 1, the central 68% (95%) credible interval for \bar{x}_{HI} is $0.49 - 0.89$ ($0.31 - 0.99$) and its posterior PDF is shown in Figure E3b. On the other hand, Model C requires a neutral universe ($\bar{x}_{\text{HI}} \sim 1$) if continuum 2 is used as input, in line with the result obtained using Model A.

Final remarks. In Extended Data Table 2 we summarize the constraints on the neutral fraction obtained by all IGM modeling methods used in this article with the two different models of the quasar intrinsic emission. In all cases we recover a large IGM neutral fraction in agreement with the analysis presented in the main manuscript. In Figure 4, we show a comparison between these three models. To be conservative, we show the constraints obtained by using continuum 1 as input as it predicts a weaker emission around the Ly α line than continuum 2 (Figure E1) and thus lower neutral fractions are needed to explain the absorption profile in J1342+0928 (Extended Data Table 2). The most conservative of the analysis is Model B, which prefers a $\bar{x}_{\text{HI}} > 0.11$ IGM at the 2σ level (see Figure E3a), being one of strongest constraints yet in the history of reionization.

Data availability. The datasets generated and analysed during this study are available from the corresponding author on reasonable request.

Code availability. We have opted not to make available the codes presented in the Method Section to model the IGM damping wing because they will be described in more detail and made available in forthcoming papers (F. Davies et al. [Model B] and M.L. Turner et al. [Model C]).

28. Suzuki, N., Tytler, D., Kirkman, D., O'Meara, J. M. & Lubin, D. Predicting QSO continua in the Ly α forest. *Astrophys. J.* **618**, 592 (2005).

29. P  ris et al. A principal component analysis of quasar UV spectra at $z \sim 3$. *Astron. Astrophys.* **530**, A50 (2011).

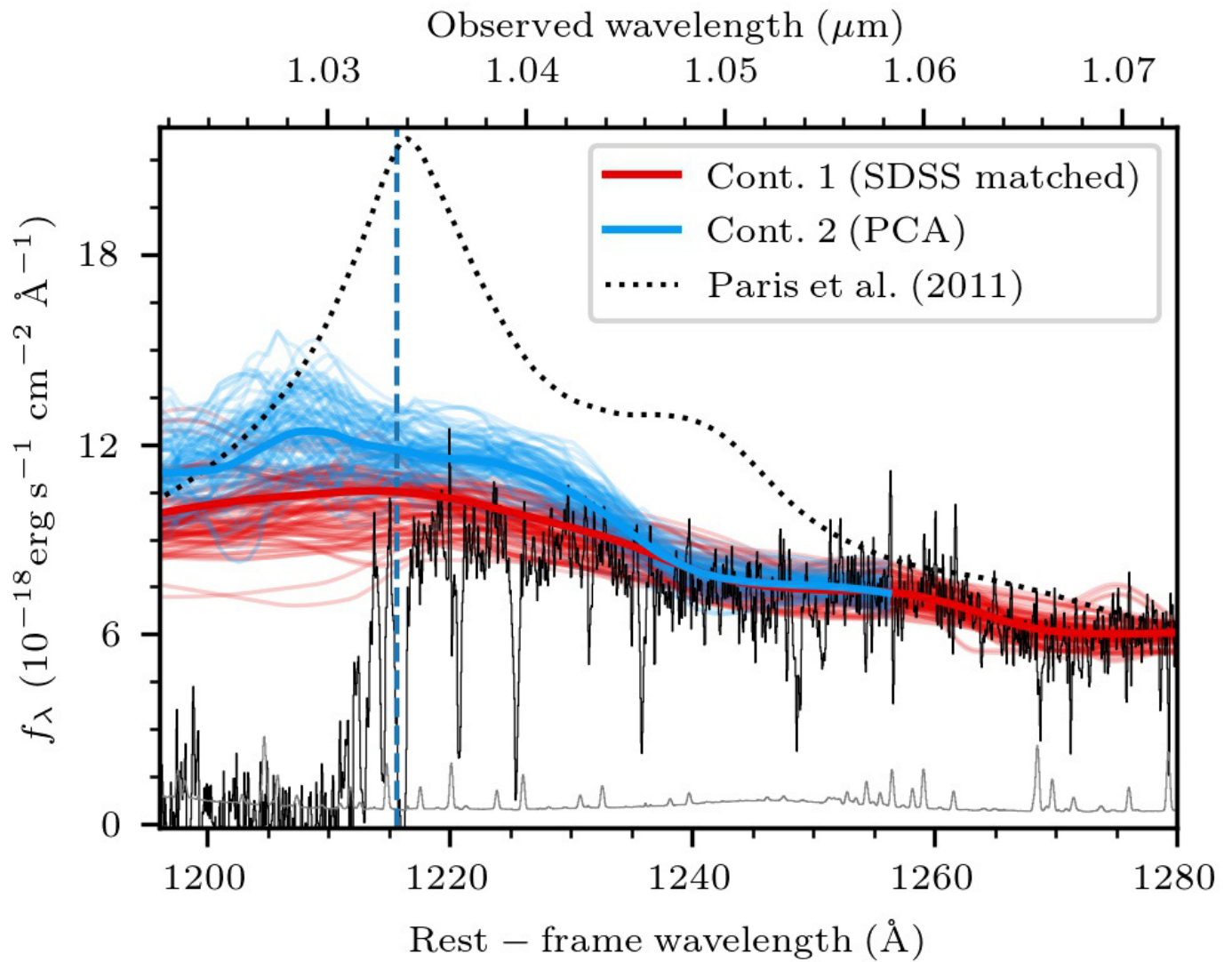
30. Alvarez, M. A. & Abel, T. Quasar HII regions during cosmic reionization. *Mon. Not. R. Astron. Soc.* **380**, L30–L34 (2007).

31. Davies, F., Furlanetto, S. & McQuinn, M. Quasar ionization from Ly α emission in an inhomogeneous intergalactic medium. *Mon. Not. R. Astron. Soc.* **457**, 3006–3023 (2016).

32. Luki   et al. The Lyman α forest in optically thin hydrodynamical simulations. *Mon. Not. R. Astron. Soc.* **446**, 3697–3724 (2015).

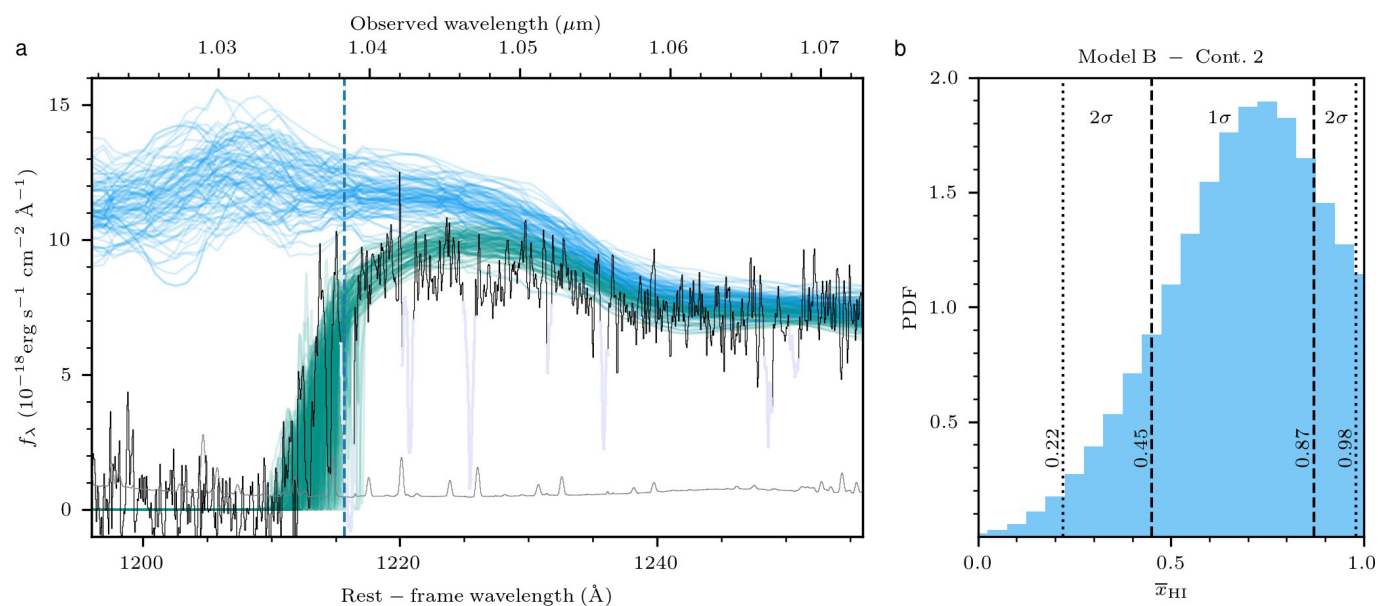
33. Mesinger, A., Furlanetto, S. & Cen, R. 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal. *Mon. Not. R. Astron. Soc.* **411**, 955–972 (2011).

34. Drovandi, C., Pettitt, A. & Lee, A. Bayesian Indirect Inference Using a Parametric Auxiliary Model. *Statistical Science*, **30**, 1, 72–95 (2015)
35. Bolton et al. 2011. How neutral is the intergalactic medium surrounding the redshift $z=7.085$ quasar ULAS J1120+0641? *Mon. Not. R. Astron. Soc.* **416**, L70–L74 (2011)
36. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: The MCMC Hammer. *Publications of the Astronomical Society of Pacific*, **125**, Issue 925, pp. 306 (2013)
37. Telfer, R. C., Zheng, W., Kriss, G. A. & Davidsen, A F. The rest-frame extreme-ultraviolet spectral properties of quasi-stellar objects. *Astrophys. J.*, **565**, 773 (2002)



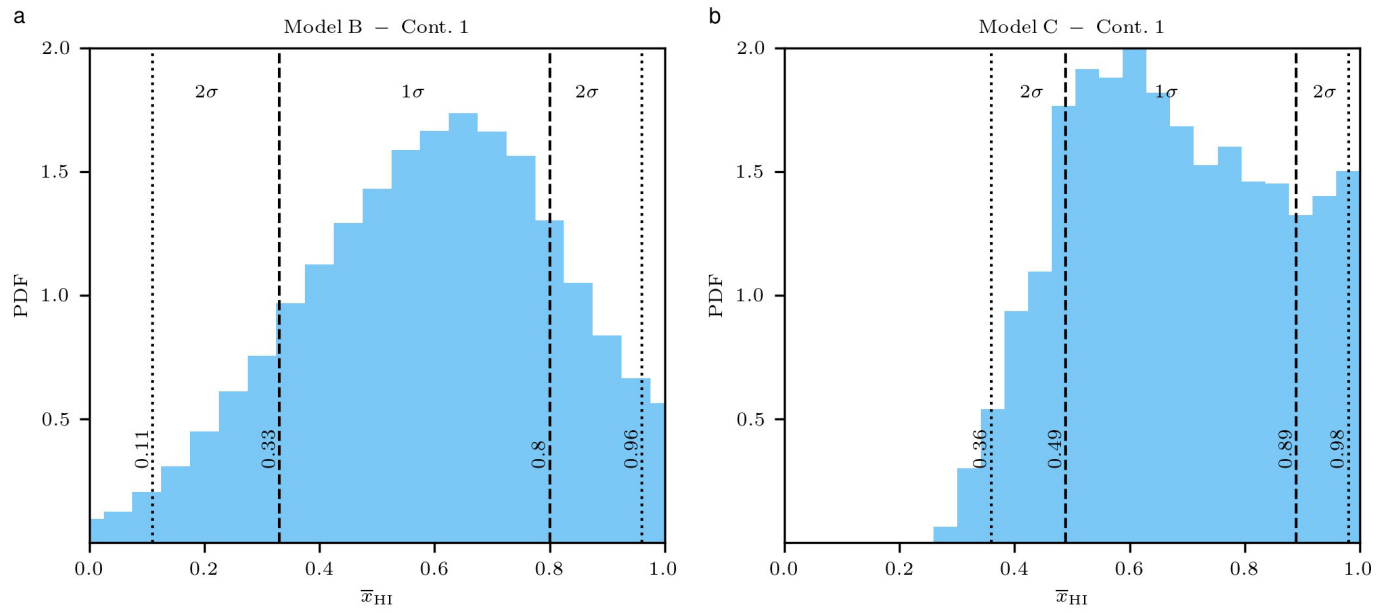
Extended Data Figure 1 | J1342+0928's intrinsic emission modeling. The red lines represent the continuum used in the main manuscript constructed by averaging SDSS DR12 quasars with similar C IV properties (EWs and blueshifts) of that observed in J1342+0928 (see Figure 3). The light blue lines are 100 random draws of PCA-reconstructed intrinsic emission as described in the Methods section. In both cases, the mean

intrinsic spectrum is shown as a thick line. The vertical dashed line shows the Ly α wavelength. The PCA-reconstructed spectrum has a stronger emission around the Ly α line than the SDSS matched reconstructed emission. The dotted line is the mean SDSS quasar from Paris et al. (2011)²⁹, which has a much stronger Ly α line than that of any of our continuum models of J1342+0928.



Extended Data Figure 2 | Model B IGM damping wing analysis with PCA continuum. **a**, Same as Figure 3b but this time showing 100 realizations of the PCA-predicted intrinsic emission (light blue) and IGM damping wing (green) model draws from the posterior PDF of Model B (see text in Methods section for details). Model B masks absorption

systems only redward of the $\text{Ly}\alpha$ line (pale gray) as this model takes into account the internal absorption in the proximity zone, which explains the larger scatter blueward of $\text{Ly}\alpha$ (dashed vertical line). **b**, The marginalized posterior PDF of \bar{x}_{HI} . The 50th percentile is $\bar{x}_{\text{HI}} = 0.68$ while the 16th – 84th (2.5th – 97.5th) percentile interval is 0.45 – 0.87 (0.22 – 0.98).



Extended Data Figure 3 | The marginalized posterior PDF of \bar{x}_{HI} using continuum 1 (SDSS-matched) as input for IGM Model B (panel a) and Model C (panel b). Model B applied to continuum 1 yields the most

conservative distribution of our analyses. Even in this case, a significantly neutral universe with $\bar{x}_{\text{HI}} > 0.11$ at the 2 σ level is preferred.

Extended Data Table 1 | Survey photometry of the quasar J1342+0928 at $z=7.54$

Survey	AB magnitudes	
DECaLS	$z_{\text{DE},3\sigma} > 23.32$	
UKIDSS	$Y = 21.47 \pm 0.19$	$J = 20.75 \pm 0.11$
	$H = 20.02 \pm 0.02$	$K = 20.03 \pm 0.12$
WISE	$W1 = 20.17 \pm 0.15$	$W2 = 20.11 \pm 0.29$

Extended Data Table 2 | Summary of the constraints on the neutral fraction $\bar{x}_{\text{H I}}$ in the surroundings of the quasar J1342+0928

Continuum	IGM Model	A – 68% (95%)	B – 68% (95%)	C – 68% (95%)
1 (SDSS-matched)		0.38 – 0.77 (0.27 – 0.94)	0.33 – 0.80 (0.11 – 0.96)	0.49 – 0.89 (0.36 – 0.98)
2 (PCA)		~ 1	0.45 – 0.87 (0.22 – 0.98)	~ 1

$\bar{x}_{\text{H I}}$ constraints from the modeling of the quasar’s Ly α damping wing with three different IGM models and two different intrinsic emission modeling. The central 68% (95%) credible intervals are reported, except when a completely neutral IGM is always the preferred solution $\bar{x}_{\text{H I}}=1$.

Orbital misalignment of the Neptune-mass exoplanet GJ 436b with the spin of its cool star

Vincent Bourrier¹, Christophe Lovis¹, Hervé Beust², David Ehrenreich¹, Gregory W. Henry³, Nicola Astudillo-Defru¹, Romain Allart¹, Xavier Bonfils², Damien Ségransan¹, Xavier Delfosse², Heather M. Cegla¹, Aurélien Wyttenbach¹, Kevin Heng⁴, Baptiste Lavie¹ & Francesco Pepe¹

The angle between the spin of a star and the orbital planes of its planets traces the history of the planetary system. Exoplanets orbiting close to cool stars are expected to be on circular, aligned orbits because of strong tidal interactions with the stellar convective envelope¹. Spin-orbit alignment can be measured when the planet transits its star, but such ground-based spectroscopic measurements are challenging for cool, slowly rotating stars². Here we report the three-dimensional characterization of the trajectory of an exoplanet around an M dwarf star, derived by mapping the spectrum of the stellar photosphere along the chord transited by the planet³. We find that the eccentric orbit of the Neptune-mass exoplanet GJ 436b is nearly perpendicular to the stellar equator. Both eccentricity and misalignment, surprising around a cool star, can result from dynamical interactions (via Kozai migration⁴) with a yet-undetected outer companion. This inward migration of GJ 436b could have triggered the atmospheric escape that now sustains its giant exosphere⁵.

Three transits of GJ 436b, which occur² every 2.64 days, were observed on 9 May 2007 (visit 1)², 18 March 2016 (visit 2) and 11 April 2016 (visit 3) with the HARPS (visit 1) and HARPS-N (visits 2 and 3) spectrographs^{6,7}. All visits cover the full transit duration, with exposure times of 300–400 s, and provide baselines of 3–8 h before or after the transit. We corrected spectra for the variability in the distribution of their flux with wavelength caused by Earth's atmosphere (Methods) before using a binary mask to calculate cross-correlation functions (CCFs) that represent an average of the spectral lines from the M dwarf host GJ 436. We introduce a double-Gaussian model to accurately fit the distinctive CCF profiles of M dwarfs (Extended Data Figs 1 and 2) and to improve the stability and precision of their derived contrast, width and radial velocity. These properties show little dispersion around their average values in each visit and are stable between the HARPS-N visits, in agreement with the low activity^{2,8} of GJ 436 (Extended Data Fig. 3).

The observed CCFs originate from starlight integrated over the disk of GJ 436 (CCF_{DI}). During the transit they are deprived of the light from the planet-occulted regions (CCF_{PO}), which we retrieve using the reloaded Rossiter–McLaughlin technique³. CCF_{DI} are shifted into the star's rest frame, then co-added and continuum-normalized outside the transit to build a master-out template CCF_{DI}^{OT} for each visit. In-transit CCF_{DI} are continuum-scaled according to the depth of the light curve derived from high-precision photometry², before subtracting them from the CCF_{DI}^{OT} to retrieve the CCF_{PO} (Methods). The local stellar line profile from the spatially resolved region of the photosphere occulted by GJ 436b along the transit chord is clearly detected in the CCF_{PO} (Fig. 1, Extended Data Fig. 4). We applied a double-Gaussian model to CCF_{PO} to derive their properties, linking the profiles of the Gaussian components in the same way as for the CCF_{DI} (Methods). We retained in our analysis all CCF_{PO} where the stellar line contrast is detected at more than 5 σ . Excluded CCF_{PO}

(Extended Data Table 1) are faint, associated with darker regions of the stellar limb that are only partially occulted by GJ 436b. The radial velocity centroids of the CCF_{PO} directly trace the velocity field of the stellar photosphere (Extended Data Fig. 5). The three series of surface radial velocities are consistent over most of the transit (even though they were obtained with two instruments over a 9-year interval) and are predominantly positive (showing that GJ 436b occults redshifted regions of the stellar disk rotating away from us and excluding an aligned system). We simultaneously fitted the three radial velocity series with the reloaded Rossiter–McLaughlin model³, using a Metropolis–Hasting Markov chain Monte Carlo algorithm⁹ and assuming a solid-body rotation for the star (Methods). The model then depends on the sky-projected obliquity λ_b (the angle between the projected angular momentum vectors of the star and of the orbit of GJ 436b) and projected rotational velocity $V_{eq} \sin i_\star$ (where i_\star is the inclination of the star spin axis relative to our line of sight). The best fit (Fig. 1, Extended Data Fig. 5) matches visits 1 and 2 well, and it yields a relatively large χ^2 of 42 for 19 degrees of freedom because three measurements in visit 3 deviate by 2.5 σ –3 σ . Excluding them yields the reduced chi-squared value $\chi^2_{red} = 1.1$ and does not change the derived properties beyond their 1 σ uncertainties (Methods), so they were retained in the final fit. Posterior probability distributions of the Markov chain Monte Carlo parameters (Extended Data Fig. 6) are well defined and yield $V_{eq} \sin i_\star = 330^{+90}_{-70}$ m s⁻¹ (>190 m s⁻¹ with 99% confidence) and $\lambda_b = 72^{+33}_{-24}^\circ$ (>30° with 99% confidence). These properties do not change beyond their 1 σ uncertainties when system parameters are varied within their error bars. The Bayesian information criterion for the best-fit solid-body model (48) is much lower than for a null velocity model (74) and an aligned model (88). The M dwarf GJ 436 is thus the coolest star across which the Rossiter–McLaughlin effect has been detected, with a highly misaligned orbit for its Neptune-mass companion (Fig. 2).

The slow rotation of GJ 436 is consistent with published upper limits^{2,10}. It yields a small amplitude of 1.3 m s⁻¹ for the classical radial velocity anomaly—much smaller than the stellar surface velocities measured with the reloaded Rossiter–McLaughlin technique—which could not be detected in earlier analyses² of visit 1. The widths of the CCF_{PO} show little dispersion around the width of the CCF_{DI}^{OT}, consistent with the non-detection of rotational broadening (Extended Data Fig. 5). The three visits show similar properties for the CCF_{PO} along the transit chord and for the CCF_{DI}^{OT}, consistent with the low activity^{11,12} of GJ 436 and stable emission at ultraviolet⁵, optical⁸ and infrared^{2,13} wavelengths. Nonetheless, small periodic variations in its visible flux⁸ and the periodic modulation we measure in the HARPS² and Keck¹⁴ chromospheric indices suggest the presence of active regions on the stellar surface.

This can be reconciled with the stability of GJ 436 emission if its spin axis is tilted⁸ so that active regions could be frequently occulted by the planet while yielding a small rotational flux modulation. Using 14 years

¹Observatoire de l'Université de Genève, 51 chemin des Maillettes, 1290 Versoix, Switzerland. ²Université Grenoble Alpes, CNRS, IPAG, F-38000 Grenoble, France. ³Center of Excellence in Information Systems, Tennessee State University, Nashville, Tennessee 37209, USA. ⁴University of Bern, Center for Space and Habitability, Sidlerstrasse 5, CH-3012 Bern, Switzerland.

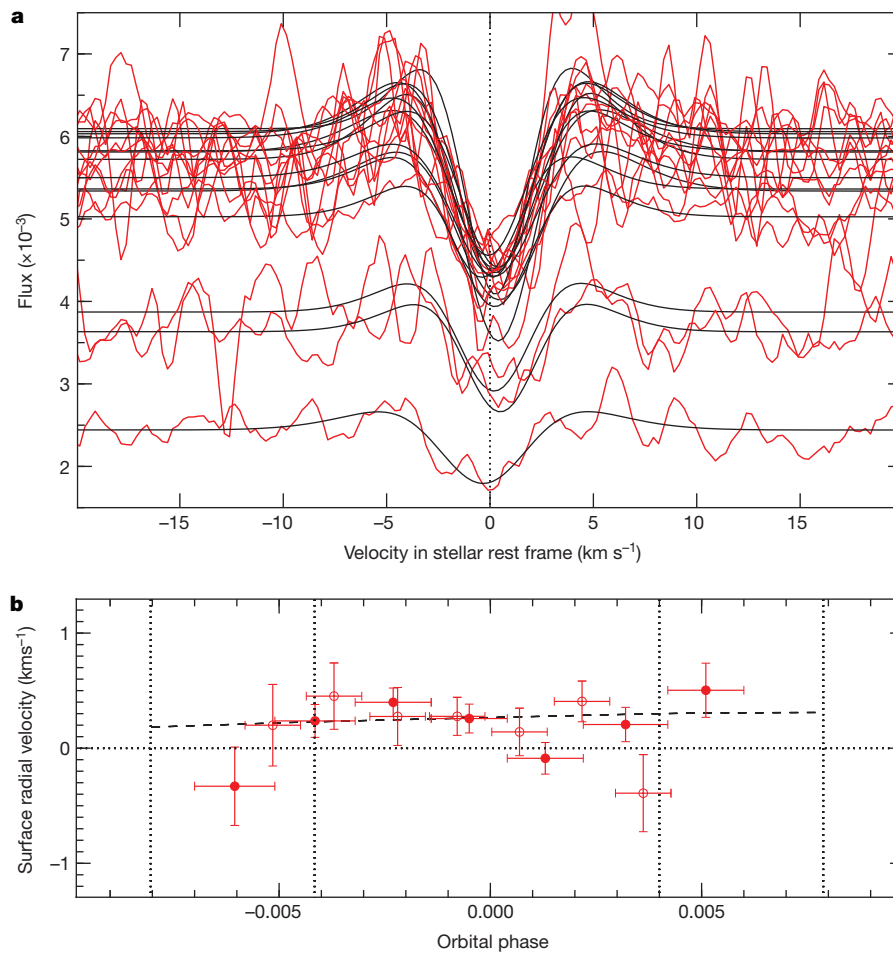


Figure 1 | Properties of the stellar photosphere along the transit chord of GJ 436b. **a**, CCF_{PO} and their double-Gaussian best fits (black lines) as a function of velocity in the stellar rest frame. Visits 2 and 3, obtained with the same instrument at similar orbital phases, were binned together. The flux level varies with limb darkening and the area occulted by the planet.

b, Intrinsic radial velocities of the stellar surface (symbols are empty for visit 1 and filled for visit 2+3) and their best-fit model (dashed line) as a function of the GJ 436b orbital phase. Dotted lines are transit contacts. Horizontal bars show the exposure durations. 1σ uncertainties are propagated from the continuum dispersion in **a**.

of ground-based differential photometry, we confirm this modulation and derive a stellar rotation period $P_{\text{rot}} = 44.09 \pm 0.08$ days, which implies that GJ 436 is older than 4 billion years (Gyr) (Methods). This value agrees well with the periods of 40.6 ± 2.2 days and 44.5 ± 4.6 days that we derive from periodograms of the $\text{H}\alpha$ and Ca II (H&K) activity indicators, respectively. Combining the stellar radius with our results for P_{rot} and $V_{\text{eq}} \sin i_{\star}$ yields $i_{\star} = 39^{\circ+13^{\circ}}_{-9^{\circ}}$ (degenerate with $i_{\star} = 141^{\circ+9^{\circ}}_{-13^{\circ}}$), confirming the tilt of the star spin axis with respect to the line of sight. By chance these degenerate values for i_{\star} yield similar distributions for the true three-dimensional (3D) obliquity of GJ 436b, which imply a nearly polar orbit with $\Psi_b = 80^{\circ+21^{\circ}}_{-18^{\circ}}$ (Fig. 2, Methods).

GJ 436b has a puzzling eccentricity² of $e_b = 0.16$: tidal interactions with the star should have circularized its orbit^{4,15} in less than about 1 Gyr, unless the internal structure of the planet results in abnormally weak tides^{4,15,16}, or a hypothetical distant companion GJ 436c perturbs its orbit. Circularization could take up to 8 Gyr if GJ 436b and GJ 436c evolved to a quasi-stationary secular fixed point in which their orbital apses are co-linear¹⁷. However, this scenario requires coplanar orbits in a specific initial configuration, which our measurement of GJ 436b's spin-orbit angle disfavours. This misalignment is unlikely to arise from scattering with a companion, as this usually occurs in young systems, and GJ 436b's orbit would have since been circularized.

It is also surprising because tides in the thick convective envelope of cool stars are expected to realign close-in planets efficiently^{1,10,18}. However, there is another outlier in the low-obliquity systems with short tidal dissipation timescales¹⁸: WASP-8b is on an eccentric¹⁹

($e = 0.3$), misaligned²⁰ ($\lambda = -143^{\circ}$) orbit that would take about as long as GJ 436b to re-align (Methods). Dynamical interactions with a massive, long-period companion have been proposed¹⁹ to explain the architecture of the WASP-8 system.

The eccentricity and obliquity⁴ of GJ 436b could originate from a similar Kozai migration induced by a possible perturber, hereafter called GJ 436c. Figure 3 shows a migration pathway that could have led to the architecture of the system in about 5 Gyr. In a first phase lasting for about 4 Gyr, GJ 436c induces strong oscillations in the eccentricity of GJ 436b and their mutual inclination, which naturally misaligns the GJ 436b orbital plane. At the onset of the second phase, the orbital distance of GJ 436b and the mutual inclination drop sharply to their present-day value. The mutual inclination keeps oscillating slightly, which results in larger oscillations of GJ 436b's 3D obliquity, consistent with the measured value. The orbit of GJ 436b, excited to a high eccentricity during the first phase, slowly circularizes and reaches the present value in about 1 Gyr. Different Kozai migrations could have led to the present architecture, and acceptable values for the initial orbit of GJ 436b, the mass and period of GJ 436c can be constrained (Methods) by combining Kozai simulations with radial velocity measurements, direct imaging, and our constraints on the age of the system (4–8 Gyr).

We illustrate this approach in Fig. 4, which shows that planetary or brown dwarf companions with masses between about 0.04 and 40 Jupiter masses and periods of 3–400 yr could have driven GJ 436b into Kozai cycles if it was initially further than about 0.2 astronomical units

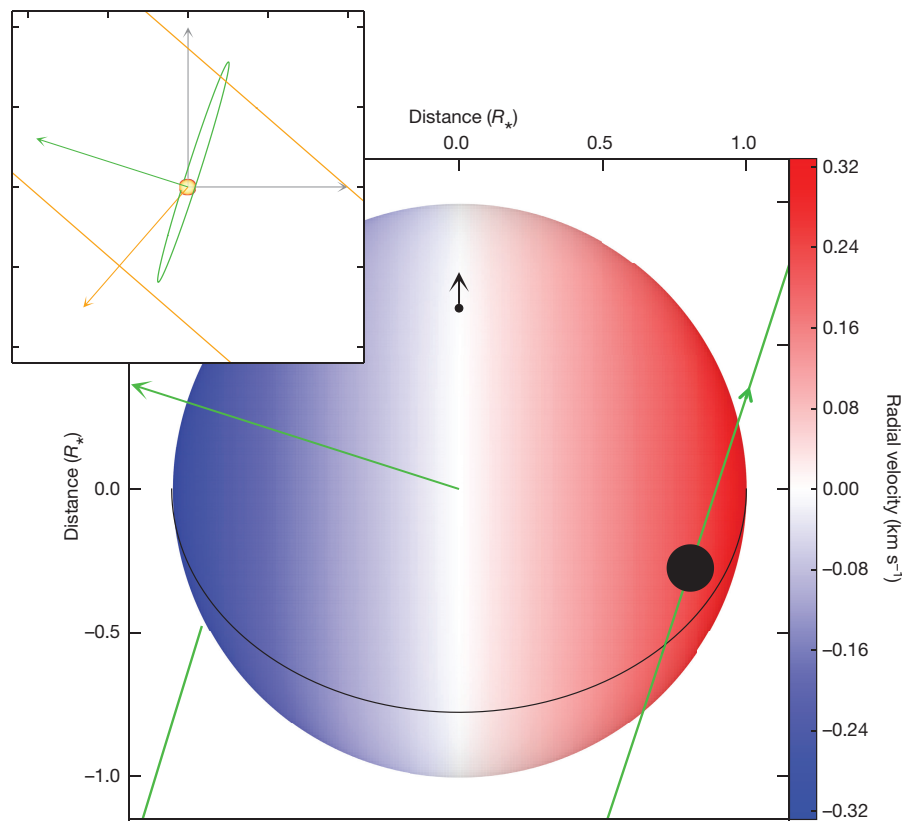


Figure 2 | Architecture of the GJ 436 system, projected on the plane of the sky. Stellar disk colour corresponds to its surface radial velocity field. The black arrow from south to north pole (visible in this configuration with $i_{\star} = 39^{\circ}$) is the inclined stellar spin. A solid black line represents the stellar equator. The orbital axis of GJ 436b (black disk) is shown as a green arrow of the same length as the half-stellar spin axis, and its orbital

trajectory as a solid green curve. The inset is a zoom of this image, in which the yellow disk represents the star, showing a possible orbit (in orange) for the perturber GJ 436c ($i_c = 89.8^{\circ}$, $\lambda_c = 139^{\circ}$, semi-major axis $a_c = 7.9$ AU; Methods). Grey axes are the sky-projected stellar spin axis and node line.

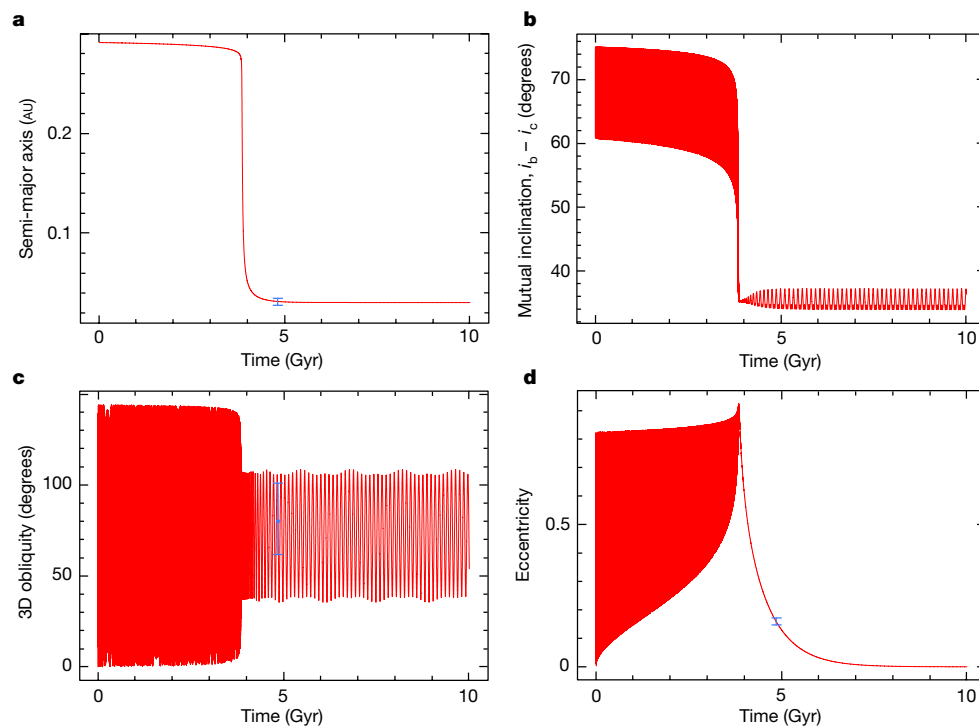


Figure 3 | Secular evolution of GJ 436b. Possible Kozai migration pathway that would have led to the present-day architecture of the GJ 436 system in about 5 Gyr, with $M_c = 0.23$ Jupiter masses the mass of GJ 436c and $a_c = 7.9$ AU its orbital distance (Methods). The semi-major axis of GJ 436b (a) and its mutual inclination with GJ 436c (b) quickly drop

once Kozai cycles end, while its eccentricity (d) slowly decreases. Low oscillations of the mutual inclination lead to larger variations of the 3D obliquity of the GJ 436b orbital plane (c). Blue points correspond to the known properties of GJ 436b.

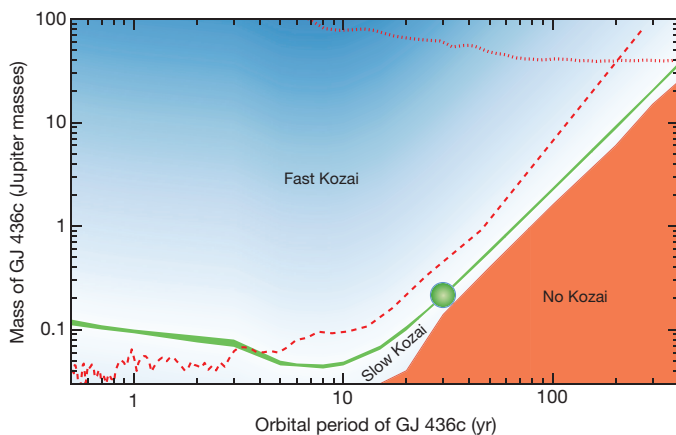


Figure 4 | Constraints on the mass and period of a putative perturber GJ 436c. The age of the system constrains the width of the green region, which delimits the properties that would have allowed GJ 436c (green disk) to drive GJ 436b to its present-day orbital configuration via Kozai migration. In the Fast Kozai region, GJ 436b would already be circularized. In the Slow Kozai region, Kozai cycles would still be ongoing. Radial velocity measurements and direct imaging exclude regions above the dashed and dotted red lines, respectively (the radial velocity curve is a limit on $M_c \sin i_c$). This diagram shows a subset of possible migrations, for the initial properties of GJ 436b (mutual inclination $i_m^0 = 85^\circ$, $a_b^0 = 0.35$ AU) and GJ 436c used in Fig. 3.

(AU) from the star. The subsequent inward migration could have altered the nature of GJ 436b, triggering the atmospheric escape that sustains the giant cloud of hydrogen trailing the planet today⁵. Meanwhile, weak tidal dissipation would have left the orbit of GJ 436c mostly unchanged over time, except for its mutual inclination with GJ 436b. By constraining its present-day value, we could determine the 3D orientation of the GJ 436c orbital plane (Methods, Fig. 2).

Since the reloaded Rossiter–McLaughlin technique directly retrieves the intrinsic stellar surface velocity, it can probe the architecture of planetary systems even around cool, slowly rotating stars. Combining the technique with next-generation infrared spectrographs (SPIRou, NIRPS) will allow for a detailed characterization of the systems discovered around M dwarfs by upcoming transit surveys (CHEOPS, TESS and PLATO). These may reveal whether GJ 436b is the exception rather than the rule.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 August; accepted 20 October 2017.

Published online 18 December 2017.

- Winn, J. N. *et al.* Hot stars with hot Jupiters have high obliquities. *Astrophys. J.* **718**, L145–L149 (2010).
- Lanotte, A. *et al.* A global analysis of Spitzer and new HARPS data confirms the loneliness and metal-richness of GJ 436 b. *Astron. Astrophys.* **572**, A73 (2014).
- Cegla, H. M. *et al.* The Rossiter–McLaughlin effect reloaded: probing the 3D spin-orbit geometry, differential stellar rotation, and the spatially-resolved stellar spectrum of star-planet systems. *Astron. Astrophys.* **588**, A127 (2016).
- Beust, H. *et al.* Dynamical evolution of the Gliese 436 planetary system. Kozai migration as a potential source for Gliese 436b's eccentricity. *Astron. Astrophys.* **545**, A88 (2012).
- Ehrenreich, D. *et al.* A giant comet-like cloud of hydrogen escaping the warm Neptune-mass exoplanet GJ 436b. *Nature* **522**, 459–461 (2015).
- Mayor, M. *et al.* Setting new standards with HARPS. *Messenger* **114**, 20–24 (2003).
- Cosentino, R. *et al.* HARPS-N: the new planet hunter at TNG. *Proc. SPIE* **8446**, 84461V (2012).

- Knutson, H. A. *et al.* A Spitzer transmission spectrum for the exoplanet GJ 436b, evidence for stellar variability, and constraints on dayside flux variations. *Astrophys. J.* **735**, 27 (2011).
- Bourrier, V. *et al.* SOPHIE velocimetry of Kepler transit candidates. XVI. Tomographic measurement of the low obliquity of KOI-12b, a warm Jupiter transiting a fast rotator. *Astron. Astrophys.* **579**, A55 (2015).
- Albrecht, S. *et al.* Obliquities of hot Jupiter host stars: evidence for tidal interactions and primordial misalignments. *Astrophys. J.* **757**, 18 (2012).
- Butler, R. P. *et al.* A Neptune-mass planet orbiting the nearby M dwarf GJ 436. *Astrophys. J.* **617**, 580–588 (2004).
- Astudillo-Defru, N. *et al.* Magnetic activity in the HARPS M dwarf sample. The rotation-activity relationship for very low-mass stars through R(HK). *Astron. Astrophys.* **600**, A13 (2017).
- Knutson, H. A. *et al.* A featureless transmission spectrum for the Neptune-mass exoplanet GJ 436b. *Nature* **505**, 66–68 (2014).
- Butler, R. P. *et al.* The LCES HIRES/Keck Precision Radial Velocity Exoplanet Survey. *Astron. J.* **153**, 208 (2017).
- Tong, X. & Zhou, J. Eccentricity modulation of a close-in planet by a companion: application to GJ 436 system. *Sci. Chin. Ser. G* **52**, 640–648 (2009).
- Mardling, R. A. On the long-term tidal evolution of GJ 436b in the presence of a resonant companion. Preprint at <https://arxiv.org/abs/0805.1928> (2008).
- Batygin, K. *et al.* A quasi-stationary solution to Gliese 436b's eccentricity. *Astrophys. J.* **699**, 23–30 (2009).
- Brothwell, R. D. A window on exoplanet dynamical histories: Rossiter-McLaughlin observations of WASP-13b and WASP-32b. *Mon. Not. R. Astron. Soc.* **440**, 3392–3401 (2014).
- Queloz, D. *et al.* WASP-8b: a retrograde transiting planet in a multiple system. *Astron. Astrophys.* **517**, L1 (2010).
- Bourrier, V. *et al.* Refined architecture of the WASP-8 system: a cautionary tale for traditional Rossiter-McLaughlin analysis. *Astron. Astrophys.* **599**, A33 (2017).

Acknowledgements This work is based on observations made with the HARPS spectrograph on the 3.6 m ESO telescope at the ESO La Silla Observatory, Chile, under GTO program ID 072.C-0488, and with the Italian Telescopio Nazionale Galileo operated on the island of La Palma by the Fundación Galileo Galilei of the INAF (Istituto Nazionale di Astrofisica) at the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias under OPTICON program 16A/049, ‘Sensing Planetary Atmospheres with Differential Echelle Spectroscopy’ (SPADES). OPTICON has received funding from the European Community's Seventh Framework Programme (FP7/2013-2016) under grant agreement number 312430. This project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme under grant agreement number 724427 (FOUR ACES). This work was carried out in the framework of the National Centre for Competence in Research ‘PlanetS’ supported by the Swiss National Science Foundation (SNSF). R.A., N.A.-D., V.B., D.E., C.L. and A.W. acknowledge the financial support of the SNSF. H.M.C. gratefully acknowledges support as a CHEOPS Fellow from the SNSF National Centre of Competence in Research ‘PlanetS’. G.W.H. acknowledges long-term support from Tennessee State University and the State of Tennessee through its Centers of Excellence programme. X.B. and X.D. acknowledge the support of CNRS/PNP (Programme national de planétologie). X.B. acknowledges funding from the European Research Council under the ERC Grant Agreement number 337591-ExTrA. We thank C. A. Watson for calculating the convective mass of GJ 436, H. Knutson for facilitating the determination of the stellar rotation period, J.-B. Delisle for discussing the system geometry, and the Telescopio Nazionale Galileo staff for the service observation.

Author Contributions V.B. coordinated the study of the GJ 436 system, performed the reduction and analysis of the transit data, interpreted the results, and wrote the paper. V.B. and D.E. proposed the original idea. D.E. developed the HARPS-N transit observation programme (led by D.E. and A.W.). V.B., H.M.C. and C.L. developed and refined the reloaded Rossiter–McLaughlin technique. H.B. performed the Kozai simulations and contributed to the interpretation. G.W.H. derived the stellar rotation period from analysis of photometry. N.A.-D. and X.D. derived the stellar rotation period from analysis of activity indices. X.B. and N.A.-D. analysed radial velocity values, and D.S. analysed direct imaging data used to constrain GJ 436c. R.A., H.C., C.L. and A.W. contributed to the analysis and interpretation of the transit data. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to V.B. (vincent.bourrier@unige.ch).

Reviewer Information Nature thanks A. C. Cameron and A. Mann for their contribution to the peer review of this work.

METHODS

Data analysis and correction of systematics. Our study is based on three transit observations of the exoplanet GJ 436b with ground-based echelle spectrographs. We obtained 77 and 71 exposures of 400 s duration with HARPS-N on 18 March and 11 April 2016, respectively, in the frame of the SPADES programme (Principal Investigator D.E.). These datasets are complemented with 44 archive exposures of 300 s duration, obtained with HARPS on 9 May 2007, which were previously used to attempt a detection of the Rossiter–McLaughlin effect². Observations were reduced with the HARPS (version 3.5) and HARPS-N (version 3.7) Data Reduction Software, yielding spectra with resolution 115,000 covering the region 380–690 nm. The reduced spectra were passed through an order-by-order cross-correlation with a M2-type mask function, weighted by the depth of the lines, to compute the cross-correlation functions (CCFs) defined in the Solar System barycentre rest frame.

The CCFs of GJ 436 display sidelobes typical of M dwarf stars (Extended Data Fig. 1). Single-Gaussian models, or Gaussian plus polynomial models limited to a portion of the CCF radial velocity range²¹, do not use the full information contained in such CCFs, which can limit the stability and the precision of their derived properties (radial velocity centroid, full-width at half-maximum, FWHM, and contrast). We pioneer a new model consisting of the sum of a Gaussian function representing the CCF continuum and sidelobes, and an inverted Gaussian function representing the CCF core. This double-Gaussian model fits well the entire CCF profile, yielding low-dispersion residuals between the CCFs and their best fit (Extended Data Fig. 1). The radial velocity centroid of the lobe component is redshifted with respect to the core component, but individual exposures show little dispersion around the average redshift in each visit (Extended Data Fig. 2). Similarly, the ratios between the amplitudes of the Gaussian components and the ratios between their FWHMs are stable in each visit. The properties of the two Gaussian components are thus tightly correlated, and we fixed the radial velocity centroid difference, the amplitude ratio, and the FWHM ratio to their average value in each night, leaving our model with only four free parameters (continuum level, radial velocity centroid, amplitude or contrast, and the FWHM of the core Gaussian component).

Earth's atmosphere induces a global variation in the flux measured during a night, leading to the loss of absolute flux levels and variations in the distribution of flux with wavelength that can be different for each exposure. This changes the relative contribution to the CCF of lines that have different width and contrast, but share similar Doppler shifts. Therefore, CCFs uncorrected for the flux unbalance show strong variations in FWHM and contrast over each night, while their radial velocities are little affected (Extended Data Fig. 3). Visit 1 is more stable than visits 2 and 3, most probably because GJ 436 culminates close to the zenith when observed with HARPS-N and thus varies strongly in elevation over the night, while it remains at similar low elevations when observed with HARPS. The reloaded Rossiter–McLaughlin technique³ relies on the comparison of the in- and out-of-transit CCFs, and therefore requires a high stability of the CCF profiles over each night. The standard correction of the flux unbalance by the HARPS and HARPS-N pipelines is not applied by default to M dwarfs because their spectra vary considerably with sub-spectral type. We thus applied a correction customized to GJ 436. For each exposure in a given night, we integrate the flux between 1/4 and 3/4 of each order in the two-dimensional extracted spectra (that is, the flux at the top of the blaze function). This yields low-resolution spectra defined as a function of the central wavelength in each order. We create a template by combining several low-resolution spectra selected for their high signal-to-noise ratio. All low-resolution spectra are normalized, divided by the template, and fitted with a sixth-order polynomial. The original two-dimensional spectra for each exposure are divided by the corresponding best-fit polynomial, before recalculating their CCFs. The corrected CCF_{DI} show a very stable contrast and FWHM (Extended Data Fig. 3), and their radial velocity values show little dispersion around the Keplerian curve calculated with GJ 436b's known orbital properties (Extended Data Table 2). In each night, some exposures have signal-to-noise ratios too low in their bluest orders to be corrected, and these are excluded from our analysis (Extended Data Table 1).

Application of the reloaded Rossiter–McLaughlin technique. CCF_{DI} are corrected for the Keplerian motion of the star² and shifted into the star rest frame using the systemic velocities derived from double-Gaussian fits to the CCF_{DI}^{OT} (9.79 km s^{−1} in all visits). Compared to Doppler tomography²², which pioneered the direct analysis of planet-induced distortions in the stellar CCFs but assumes constant photospheric line profiles, the reloaded Rossiter–McLaughlin technique enables a cleaner isolation of the planet's velocity–space trajectory across the star through the analysis of the local CCF_{PO} obtained by subtracting the in-transit CCF_{DI} from the CCF_{DI}^{OT}. Since the absolute flux levels of the CCF_{DI} are lost, the CCF_{DI} were calibrated photometrically using the GJ 436b transit light curve calculated with the *batman* package²³. We used nonlinear limb-darkening coefficients

derived from the transit photometry of GJ 436b in a visible band²⁴ covering most of the HARPS and HARPS-N spectral range (Extended Data Table 2). CCF_{PO} are assigned flux errors set to the standard deviation in the flat region of their continuum. Since CCFs are oversampled by the instrument pipelines (steps of 0.25 km s^{−1} for a pixel width of 0.82 km s^{−1}), we measured the standard deviation after removing three in four points. Uncertainties on the parameters derived from the double-Gaussian fits to the CCF_{PO} are 1 σ statistical errors from a Levenberg–Marquardt least-squares minimization. We assumed that all CCFs of GJ 436 share similar double-Gaussian profiles, that is, that the difference between the radial velocity centroids of the lobe and core components of the CCF_{PO}, the ratio between their amplitudes, and the ratio between their widths, were set to the average values derived from the fits to the CCF_{DI} in each visit. This assumption was validated a posteriori by the good fit of this model to the CCF_{PO}, with no spurious features found in the residuals.

While the shape of the transit light curve must be known to apply the reloaded Rossiter–McLaughlin technique, the orbital properties and ephemeris of GJ 436b could potentially be derived from the analysis of the surface radial velocity values. However, these properties are determined more precisely through photometry and velocimetry than through analysis of the Rossiter–McLaughlin effect⁹, and were thus fixed to the values in Extended Data Table 2. Nonetheless, we varied each of these parameters within their 1 σ uncertainties and confirmed that the associated surface radial velocities never deviated beyond the 1 σ uncertainties of the nominal values in Fig. 1.

Analysis of the stellar surface velocity field. Under the assumption of solid-body rotation (reasonable for mid-M dwarfs²⁵), V_{eq} and i_{\star} are degenerate because analysis of the surface radial velocities alone does not allow the determination of the stellar latitudes transited by the planet. We thus fitted λ_b and $V_{\text{eq}}\sin i_{\star}$ with the reloaded Rossiter–McLaughlin model³ using uniform priors in a custom-made Markov chain Monte Carlo algorithm⁹. We applied an adaptive principal component analysis so that step jumps take place in an uncorrelated space, which samples the posterior distributions better. We analysed the system with multiple chains, starting at random points in the parameter space. We checked that all chains converged to the same solution, thinned them using the maximum correlation length of the parameters, and merged them to obtain posterior distributions with a sufficient number of independent samples. The best-fit values for the model parameters are set to the medians of the posterior probability distributions and their 1 σ uncertainties are evaluated by taking limits at 34.15% on either side of the median (Extended Data Fig. 6).

GJ 436 passed close to the zenith in visits 2 and 3, which can lead to tracking issues with the HARPS-N telescope (Telescopio Nazionale Galileo) owing to its altazimuth mount. This occurred much earlier than the transit in visit 2 (near phase −0.049), with no apparent negative effects on our results (Extended Data Fig. 3). In visit 3, Telescopio Nazionale Galileo staff astronomers reported tracking issues with exposures at phases 0.0031 and 0.0052. GJ 436 culminated just after phase 0.0031 (elevation 87.85°), and exposures on both sides were also taken close to the zenith with elevations of 87.49° (phase 0.0014) and 87.17° (phase 0.0052). Thus, fibre injection issues might have affected the three last in-transit exposures in visit 3 (Extended Data Fig. 5), which could explain the two radial velocity deviations observed at phases 0.0014 and 0.0031. However, the radial velocity of the last exposure at phase 0.0052 is consistent with the best-fit model and with the other visits, and the contrast and FWHM of these three last in-transit exposures show no deviations compared to the other visits. Finally, the largest of the three radial velocity deviations in visit 3 comes from the first CCF_{PO} during ingress, which is faint and might thus yield less accurate measurement. Since the origin of these deviations is not clear, and they do not substantially influence the derived best-fit model, we kept them in our analysis.

Rotation period and age of GJ 436. We observed GJ 436 during 14 seasons between November 2003 and May 2017 with the T12 0.80-m Automatic Photoelectric Telescope at Fairborn Observatory in Arizona²⁶. This yielded 1,986 measurements in the Strömgren *b* and *y* photometric pass bands, combined into a single pass band to improve the precision (about 1.5–2.0 mmag for a single observation). We computed differential magnitudes of GJ 436 versus the mean brightness of two comparison stars (HD102555 and HD103676), which were constant to within 1 mmag during all observing seasons. Extended Data Fig. 7 shows the nightly differential magnitudes after observations in the transit window were removed. Observations were corrected for long-term variations and normalized so that each observing season has the same mean, yielding an overall dispersion of 4.1 mmag. We performed a frequency analysis based on least-squares sine fits with trial periods between 1 and 100 days. The goodness of fit at each period is measured as the reduction factor in the variance of the data, yielding a clear detection at 44.09 ± 0.08 days. The uncertainty is derived from the FWHM of the peak associated with this photometric period, which we interpret as the stellar rotation period P_{rot} made evident by rotational modulation in the visibility of surface

starspots (Extended Data Fig. 7). Five out of the 14 individual seasons show definitive periodic variations in agreement with P_{rot} , ranging from 41.7 days to 46.6 days with a weighted mean of 44.44 ± 0.30 days.

We used our measurement of P_{rot} to constrain the age of GJ 436, estimated at 3.7 ± 3.9 Gyr by its observed effective temperature and bolometric flux²⁷. Observations of cool stars in open clusters show that stellar rotation periods increase with redder colour (lower mass). Stars in the 2.5-Gyr-old cluster NGC 6819²⁸ have a lower spin-down rate for $B - V > 0.65$, the period increasing from 19 days to 23 days when $B - V$ increases from 0.65 to 0.88. Since this rate is not expected to increase with lower masses, we can extrapolate to see that GJ 436 ($B - V = 1.45$) would rotate in a maximum of 33 days if it were 2.5 Gyr old, showing that it is in fact much older. Cool stars in the open cluster M67, aged 4.2 Gyr, show a similar flattening of the spin-down rate²⁹ for $B - V > 0.65$, the period increasing from about 25–30 days to 30–35 days when $B - V$ increases from 0.65 to 1. With $P_{\text{rot}} = 44$ days the age of GJ 436 is likely to be close to 4–5 Gyr, and we consider 4–8 Gyr to be a conservative range.

Inclination of the star spin axis and 3D obliquity of GJ 436b. We combined our measurement of the period and the stellar projected rotational velocity to derive the inclination of the star spin axis $i_{\star} = \arcsin[P_{\text{rot}} V_{\text{eq}} \sin i_{\star} / (2\pi R_{\star})]$, with R_{\star} the stellar radius. It is then possible to determine the 3D obliquity between the orbital plane normal to GJ 436b and the spin axis of the star, $\Psi_b = \arccos(\sin i_{\star} \cos \lambda_b \sin i_b + \cos i_{\star} \cos i_b)$, with i_b the orbital inclination of GJ 436b). To determine best-fit values and uncertainties for i_{\star} and Ψ_b we randomly sampled their probability distributions, assuming a Gaussian distribution for P_{rot} and using the Markov chain Monte Carlo probability distributions obtained for $V_{\text{eq}} \sin i_{\star}$ and λ_b . There remains a degeneracy between the star spin axis pointing towards or away from the observer, yielding $i_{\star} = 39^{+13}_{-9}$ or $i_{\star} = 141^{+9}_{-13}$. Nonetheless, because of the high projected obliquity the corresponding values for Ψ_b (77^{+20}_{-15} or 82^{+19}_{-15}) are compatible with each other. We consider their average, 80^{+21}_{-18} , as the 3D obliquity of the system.

Tidal dissipation timescale of GJ 436b. We placed the GJ 436 system in figure 4 of ref. 18, which shows obliquity measurements as a function of $\tau = (M_b/M_{\text{conv}})^{(-1/3)}(a/R_{\star})$, a quantity proportional to the mass of the stellar convective envelope (M_{conv}) and to the scaled distance to the star (a_b/R_{\star}), and thus to the tidal dissipation timescale (where M_b is the mass of GJ 436b). We derived $M_{\text{conv}} \approx 0.146$ for GJ 436, using the EZ_WEB stellar evolution code (<http://www.astro.wisc.edu/~townsend/static.php?ref=e-z-web>; this result is largely insensitive to the age of the star and its initial mass). Figure 4 in ref. 18 shows that systems with short tidal dissipation timescales ($\tau \leq 700$) are preferentially aligned ($\lambda \leq 20^\circ$). The only two outliers in this distribution of low-obliquity systems are GJ 436 ($\tau \approx 180$, $\lambda_b = 72^\circ$) and WASP-8 ($\tau \approx 240$, $\lambda = 143^\circ$).

Kozai migration of GJ 436b. The Kozai migration of GJ 436b was presented and simulated with an N -body + tides code in ref. 4. We show in Fig. 3 a possible evolution based on our new constraints on the system. The semi-major axis of GJ 436b had to be initially larger than it is today, to prevent tidal effects circularizing its orbit too fast. During a first phase GJ 436c induced strong oscillations of the eccentricity and inclination of GJ 436b. At peak eccentricity, inclination and periastron are minimal, and tidal friction slowly decreases the semi-major axis. The bottom eccentricity of the Kozai cycles gradually increases, until it reaches the peak eccentricity and the cycles stop. The orbital distance of GJ 436b and its mutual inclination with GJ 436c then drop sharply because of tides, while the eccentricity of GJ 436b (excited to high values at the onset of the second phase) decreases slowly to its present value. Kozai cycles in the first phase misaligned the orbit of GJ 436b (initially assumed to be within the stellar equatorial plane), leading to strong oscillations of its 3D obliquity. During the second phase the orbit of GJ 436b remains misaligned, and its 3D obliquity keeps oscillating at a slower rate between about 40° and 105° , in agreement with the measured Ψ_b .

Kozai migration primarily depends on the mass M_c and semi-major axis a_c of the perturber GJ 436c, the initial semi-major axis a_b^0 of GJ 436b, and the parameter $h^0 = |\cos i_m^0| \sqrt{1 - (e_b^0)^2}$ (with e_b^0 the initial eccentricity of GJ 436b, and i_m^0 its mutual inclination with GJ 436c). Our goal is not to explore the full parameter space, but to show that Kozai migration can explain the architecture of the system with no need for an abnormal tidal dissipation factor for GJ 436b, which was thus set to a Neptune-like value of 10^5 (ref. 4). We used the age of the system (4–8 Gyr) to constrain the transition time t_{tr} between the two phases of the Kozai evolution. This transition time delimits three regions in the (a_c, M_c) plane (Fig. 4): the ‘fast Kozai’ region ($t_{\text{tr}} < 4$ Gyr), excluded because GJ 436b would be circularized today; the ‘slow Kozai’ region ($t_{\text{tr}} > 8$ Gyr), excluded because the Kozai cycles would still be ongoing today; and the ‘convenient’ region, which allows GJ 436b to be in the later stages of the second phase within the age range of the system. For a given set of initial properties (a_b^0, h^0), the convenient region thus defines the acceptable values of (a_c, M_c) for GJ 436c, upon which we can further place upper limits derived from radial velocity measurements and adaptive optics imaging (see below).

We find that the present system architecture can be explained if GJ 436b initially satisfied $a_b^0 \geq 0.2$ AU and $h^0 \leq 0.17$ (that is, $i_m^0 \geq 80^\circ$ for small e_b^0). In that case, the Kozai migration could have been driven by perturbers with masses between about 0.04 and 40 Jupiter masses and periods between about 3 yr and 400 yr (Fig. 4). We note that other migration pathways exist, different initial conditions for GJ 436b shifting the width and position of the convenient region in the (a_c, M_c) plane. Future radial velocity values and direct imaging measurements will refine the constraints on these properties.

Conditions on GJ 436c orbital trajectory. The mutual inclination between the orbital planes of GJ 436b and GJ 436c satisfies $\cos i_m = \cos i_b \cos i_c + \cos \Omega \sin i_b \sin i_c$, with i_b and i_c the inclinations of the orbital planes, and $\Omega = \omega_c - \omega_b$ the difference between the longitudes of their ascending nodes. Since i_c is known to a high precision, the values satisfying this relation follow the 3D surface shown in Extended Data Fig. 8. If the mutual inclination i_m is known, this relation reduces to an oval ring in the (Ω, i_c) plane. Furthermore, if we take the sky-projected node line of the star as a reference for the longitude of the ascending node ω , the sky-projected obliquity of an orbiting body satisfies $\lambda = \omega$ or $\lambda = \omega - 180^\circ$. It is then possible to constrain the alignment of GJ 436c with $\lambda_c = \lambda_b + \Omega$ or $\lambda_c = \lambda_b + \Omega - 180^\circ$. Constraints on the mutual inclination would thus allow a full determination of GJ 436c orbital trajectory. This will require a complete exploration of Kozai migration pathways, which is beyond the scope of this study. Here, we illustrate this point with the scenario shown in Fig. 3, where the mutual inclination oscillates between 66° and 68° and constrains $|i_c - 90^\circ| < 71^\circ$, $|\Omega| < 68^\circ$, and λ_c within $[-20^\circ, 173^\circ]$ or $[-200^\circ, -6^\circ]$. A possible trajectory for GJ 436c is shown in Fig. 2, where we selected $i_m = 67^\circ$ and $i_c = 89.8^\circ$, yielding $\Omega = 67^\circ$ and $\lambda_c = 139^\circ$. The semi-major axis $a_c = 7.9$ AU was derived from Fig. 4.

We note that two transiting Earth-sized companions have been postulated in the GJ 436 system³⁰, on shorter and larger orbits than GJ 436b. However, they were not confirmed by later analyses², and could not have driven the Kozai migration of GJ 436b given the results of our simulations⁴ (Extended Data Fig. 5) and the constraints on their properties derived from radial velocity measurements² and transit studies^{31–33}.

Constraints on GJ 436c from radial velocities and direct imaging. We derived conservative detection limits on $M_c \sin i_c$ from the residuals of the HARPS² and Keck¹⁴ radial velocity time series using the same approach as in ref. 2. Perturbers above the red line in Fig. 4 are excluded for a given period with a 99% confidence level. We note that the constraint on the true mass of GJ 436c depends on its orbital inclination, which could be derived as explained above.

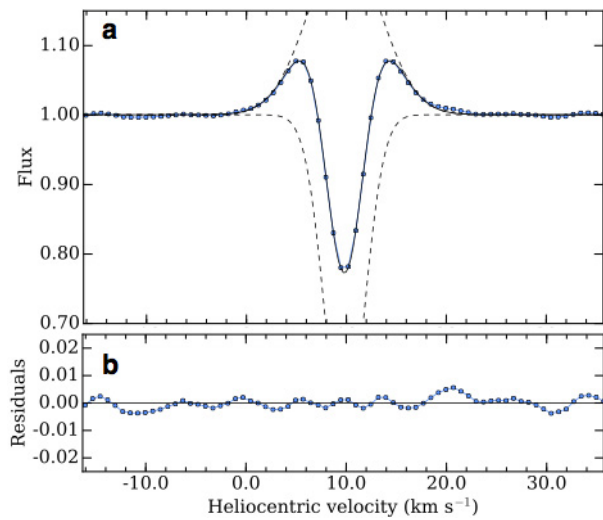
We retrieved from the ESO archive (programme 081.C-0430; Principal Investigator D. Apai) publicly available high-contrast imaging data of GJ 436 taken at the Very Large Telescope with the Nasmyth Adaptive Optics System (NAOS) Near-Infrared Imager and Spectrograph (CONICA) instrument. The data were taken in April 2008 in the L' band, using the field tracking mode of NACO, no coronagraph, and no image saturation. We used the Geneva High Contrast Imaging Data Reduction Pipeline³⁴ to reduce the data and compute the $\delta L'$ band detection limits. Since no L' photometry could be found in the literature for GJ 436, we estimated it using near-infrared photometry and stellar evolutionary models. We used the low-mass star models of ref. 35 at an age of 5 Gyr and with solar metallicity, apparent magnitudes of the 2MASS J, H and K_s bands and the Hipparcos parallax. We obtained a mid-infrared magnitude estimate $L' = 5.78 \pm 0.03$ for GJ 436, which corresponds to a mass of $M_{\star} = 0.46 M_{\odot}$ and an effective temperature of $T_{\text{eff}} = 3,610$ K, in good agreement with the spectroscopic analysis²⁷ (Extended Data Table 2). The absolute L' -band detection limits as a function of the projected separation are obtained by combining the results of the NACO images and the magnitude estimate of GJ 436, while the conversion into the companion's mass detection limits is done using the evolutionary models of ref. 36 for cool brown dwarfs. Figure 4 shows that the presence of massive brown dwarfs ($M > 40$ Jupiter masses) at long periods ($P > 90$ yr) is ruled out.

Code availability. We have opted not to make available the codes used for the data extraction and analysis as they are currently an important asset of the researchers' tool kits.

Data availability. All spectra used in this study are publicly available on the ESO archive (HARPS; http://archive.eso.org/eso/eso_archive_main.html) and on the Telescopio Nazionale Galileo archive (HARPS-N; <http://archives.ia2.inaf.it/tng/>). Source Data for Fig. 1 are available online. The other data sets generated and analysed during the present study are available from the corresponding author on reasonable request.

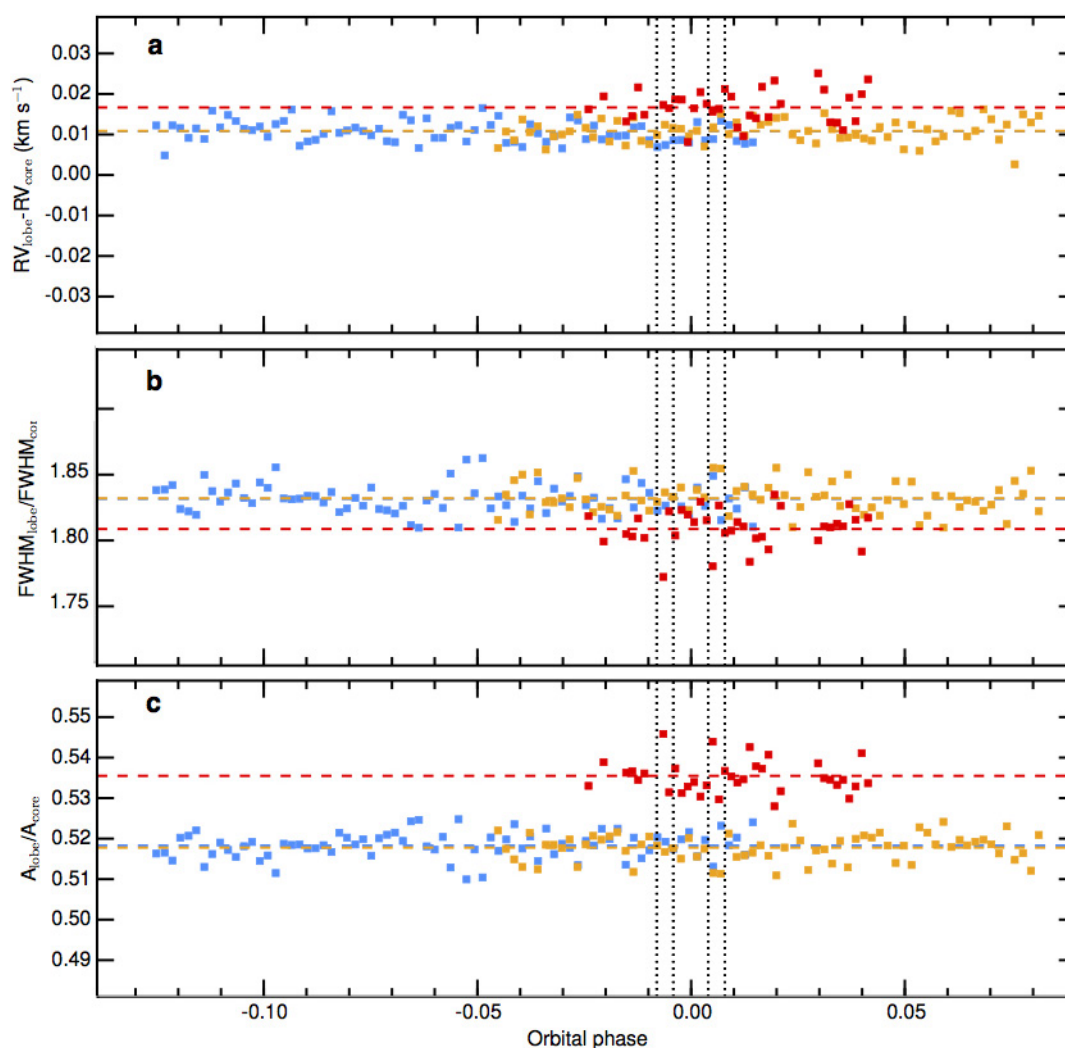
21. Suárez Mascareño, A. *et al.* Rotation periods of late-type dwarf stars from time series high-resolution spectroscopy of chromospheric indicators. *Mon. Not. R. Astron. Soc.* **452**, 2745–2756 (2015).
22. Collier Cameron, A. *et al.* Line-profile tomography of exoplanet transits—I. The Doppler shadow of HD 189733b. *Mon. Not. R. Astron. Soc.* **403**, 151–158 (2010).

23. Kreidberg, L. batman: BAsic Transit Model cAlculationN in Python. *Publ. Astron. Soc. Pacif.* **127**, 1161 (2015).
24. Bean, J. L. *et al.* EXOFAST: A Hubble Space Telescope transit light curve for GJ 436b. *Astron. Astrophys.* **486**, 1039 (2008).
25. Morin, J. *et al.* Large-scale magnetic topologies of mid M dwarfs. *Mon. Not. R. Astron. Soc.* **390**, 567–581 (2008).
26. Henry, G. W. Techniques for automated high-precision photometry of Sun-like stars. *Publ. Astron. Soc. Pacif.* **111**, 845 (1999).
27. Mann, A. W. *et al.* How to constrain your M dwarf: measuring effective temperature, bolometric luminosity, mass, and radius. *Astrophys. J.* **804**, 64 (2015).
28. Meibom, S. *et al.* A spin-down clock for cool stars from observations of a 2.5-billion-year-old cluster. *Nature* **517**, 589–591 (2015).
29. Barnes, S. A. *et al.* Rotation periods for cool stars in the 4 Gyr old open cluster M67, the solar-stellar connection, and the applicability of gyrochronology to at least solar age. *Astrophys. J.* **823**, 16 (2016).
30. Stevenson, K. B. *et al.* Two nearby sub-Earth-sized exoplanet candidates in the GJ 436 system. *Astrophys. J.* **755**, 9 (2012).
31. Alonso, R. *et al.* Limits to the planet candidate GJ 436c. *Astron. Astrophys.* **487**, L5 (2008).
32. Ballard, S. *et al.* A search for additional planets in the NASA EPOXI observations of the exoplanet system GJ 436. *Astrophys. J.* **716**, 1047 (2010).
33. Maciejewski, G. *et al.* On the GJ 436 planetary system. *Acta Astron.* **64**, 323 (2014).
34. Hagelberg, J. *et al.* The Geneva reduction and analysis pipeline for high-contrast imaging of planetary companions. *Mon. Not. R. Astron. Soc.* **455**, 2178–2186 (2016).
35. Baraffe, I. *et al.* New evolutionary models for pre-main sequence and main sequence low-mass stars down to the hydrogen-burning limit. *Astron. Astrophys.* **577**, A42 (2015).
36. Baraffe, I. *et al.* Evolutionary models for cool brown dwarfs and extrasolar giant planets. The case of HD 209458. *Astron. Astrophys.* **402**, 701 (2003).



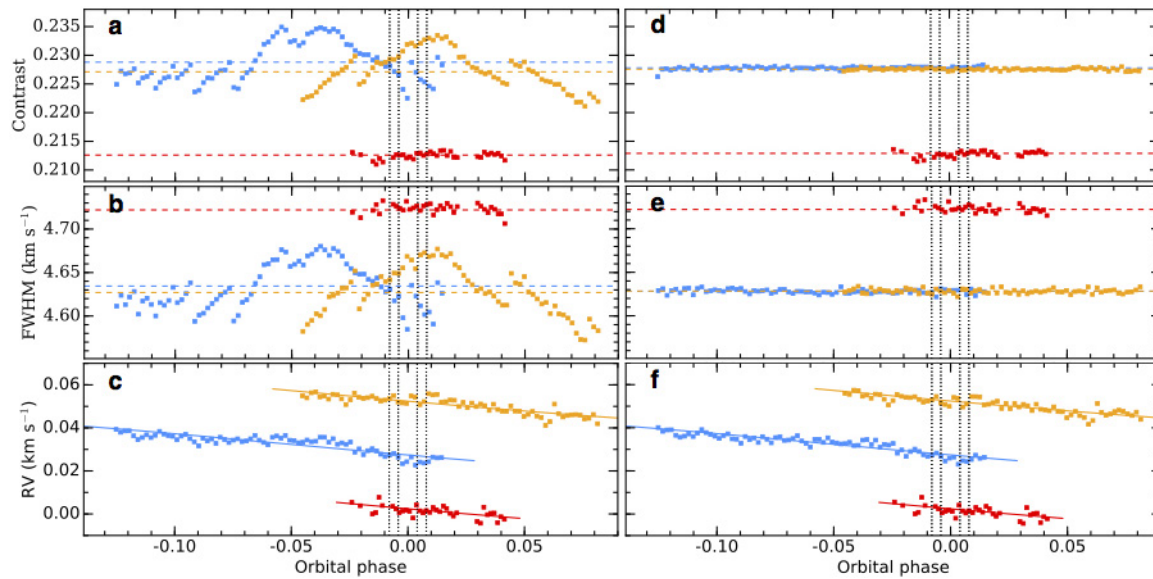
Extended Data Figure 1 | Observed and modelled CCF of GJ 436.

a, Typical HARPS-N CCF of GJ 436 (blue points), fitted with a double-Gaussian model (solid black line). This model is the combination of a Gaussian profile for the CCF continuum and lobes plus an inverted Gaussian profile for the CCF core (individual components are plotted as dashed black lines). **b**, Residuals between the observed CCF and its best fit.



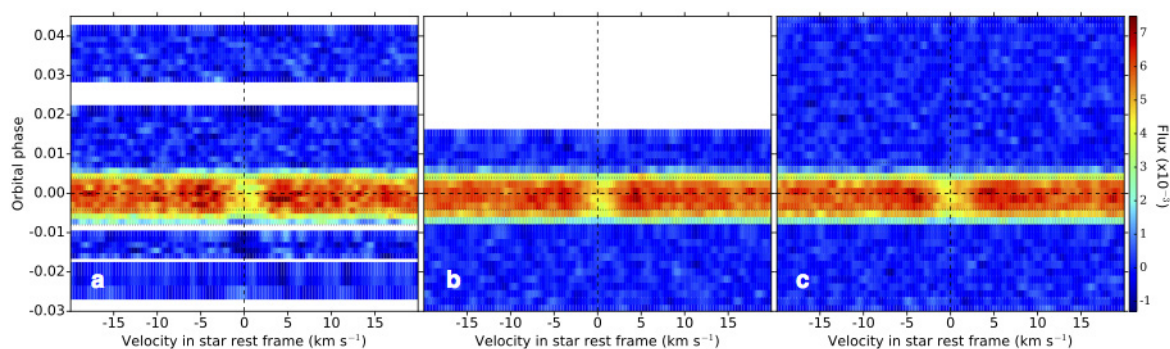
Extended Data Figure 2 | Comparison between the properties of the lobe and core Gaussian components of the CCF model. The panels show the difference between the radial velocity (RV) centroids of the lobe and core components (a), the ratio between their FWHMs (b), and the ratio

between their amplitudes (c), as a function of GJ 436b's orbital phase for each exposure in visit 1 (red), visit 2 (blue) and visit 3 (orange). There is little dispersion of these values around their average in each visit, shown as dashed horizontal lines. Vertical dotted lines are the transit contacts.



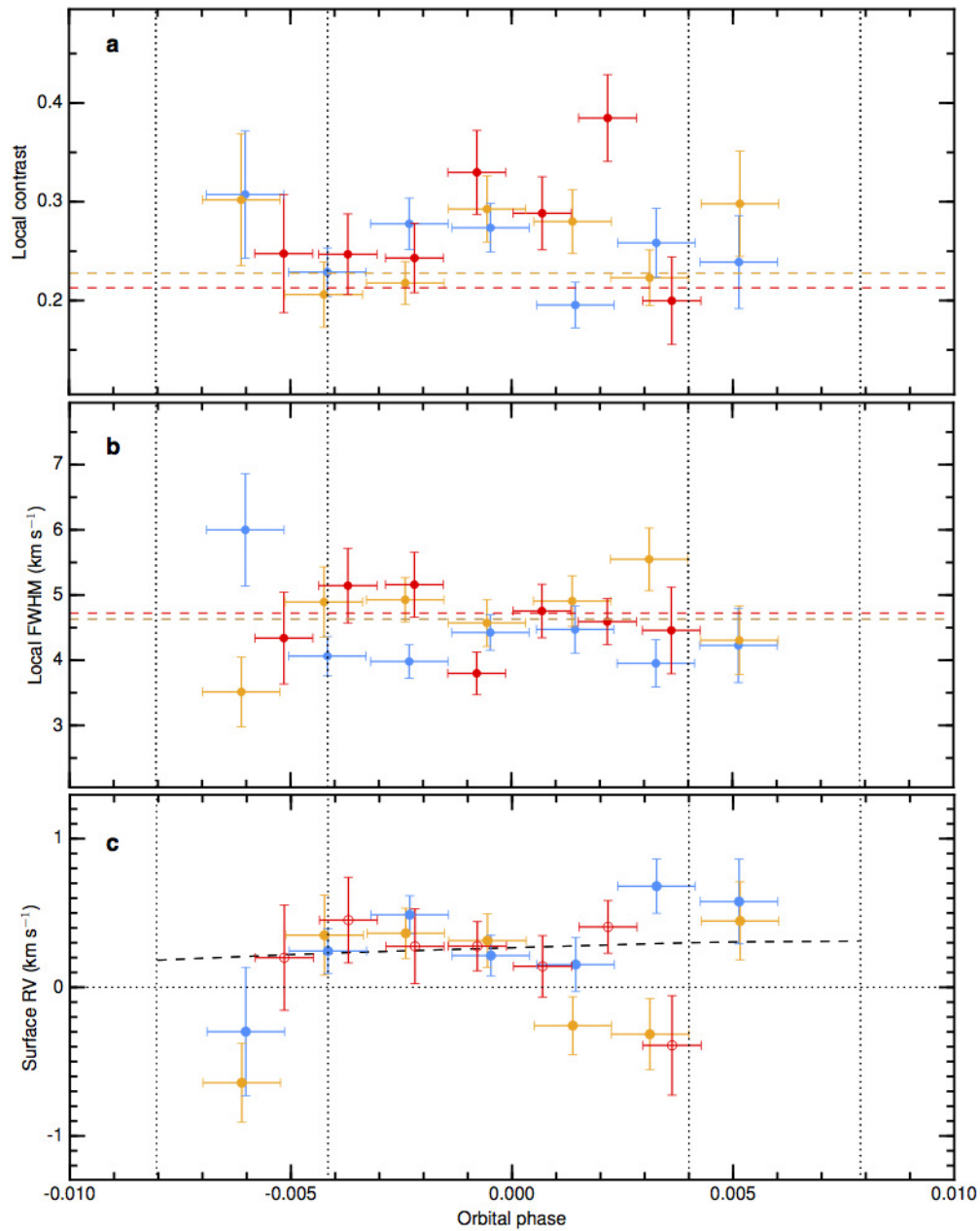
Extended Data Figure 3 | Correction for the effects of Earth's atmosphere. Properties derived from the double-Gaussian fits to the CCF_{DI} are shown before correction (a–c) and after correction of the flux distribution (d–f), as a function of GJ 436b's orbital phase. The contrast of the CCF_{DI} is shown in a and d, their FWHMs in b and e, and their radial

velocities in c and f. Radial velocities are relative to the systemic velocity in each visit, and have been offset by 25 m s^{-1} . They are overplotted with the expected Keplerian radial velocity curve. Visits 1, 2 and 3 are coloured in red, blue and orange, respectively. Vertical dotted lines are the transit contacts; horizontal dashed lines show the average values in each visit.



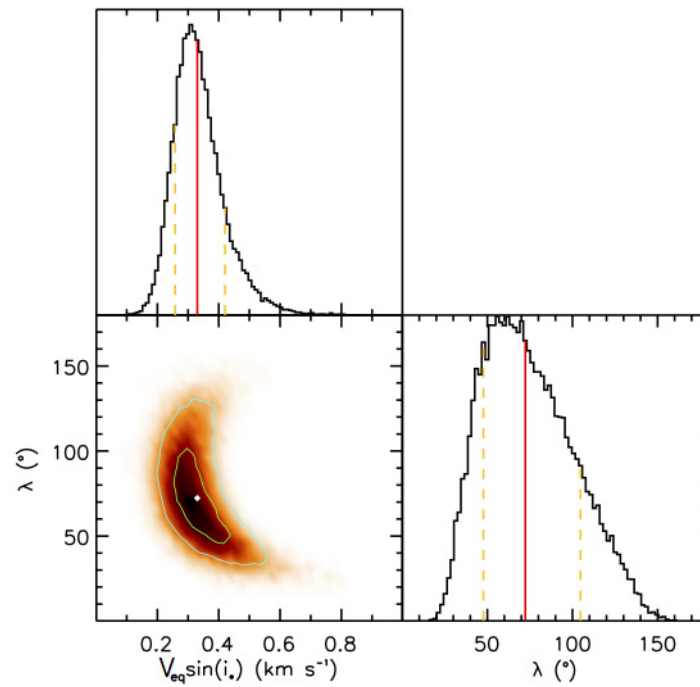
Extended Data Figure 4 | Maps of the residuals between the scaled CCF_{D1} and the $\text{CCF}_{\text{D1}}^{\text{OT}}$. Residuals are coloured as a function of their flux, and plotted as a function of radial velocity in the stellar rest frame (in abscissa) and orbital phase (in ordinate) for visit 1 (a), visit 2 (b) and visit 3 (c). The vertical and horizontal dashed black lines indicate the

mid-transit time and stellar rest velocity, respectively. In-transit residuals correspond to the CCF_{PO} , and show the average stellar line profile (recognizable by a lower flux in the CCF_{PO} cores) from the regions occulted by GJ 436b across the stellar disk. Out-of-transit residuals show little dispersion in all visits, consistent with the low activity of the host star.



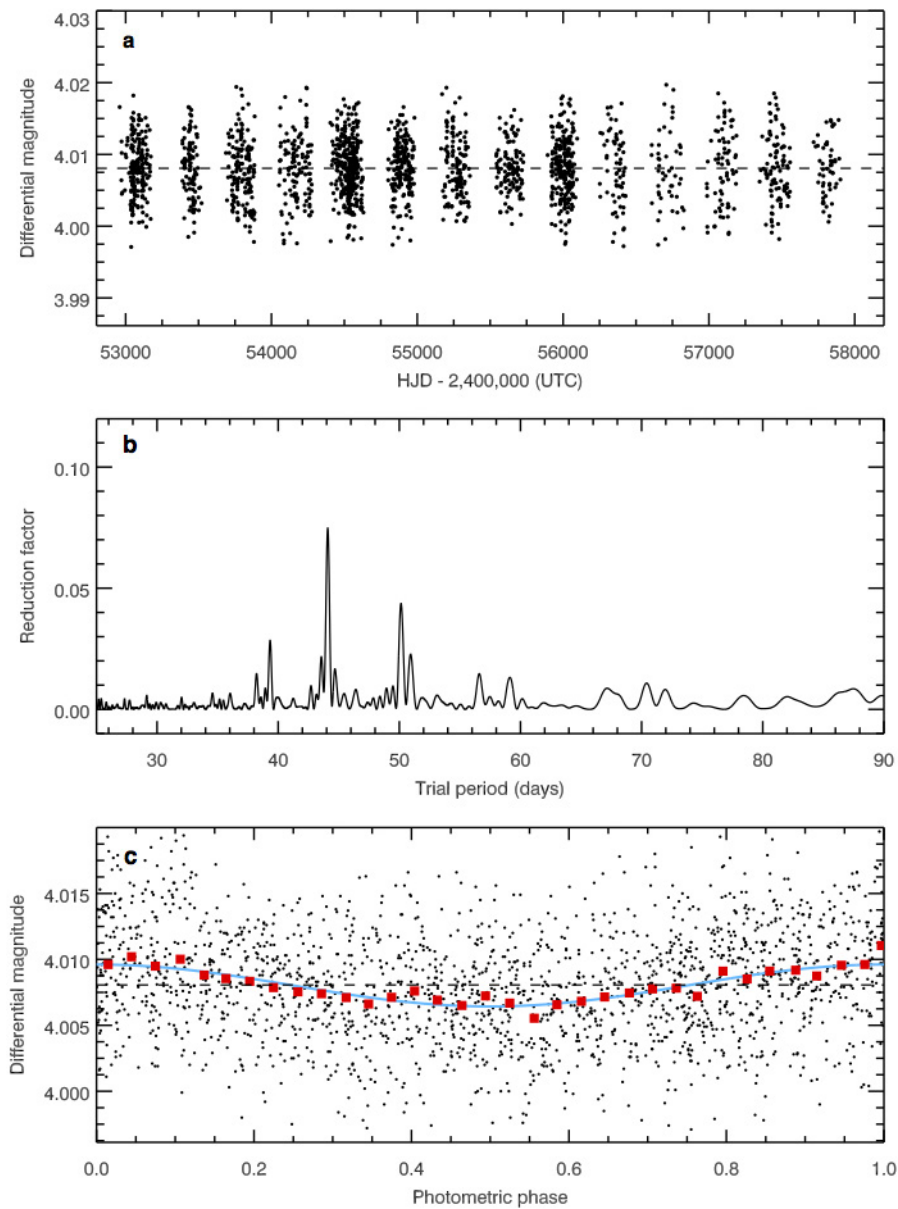
Extended Data Figure 5 | Properties of the CCF_{PO} as a function of GJ 436b orbital phase. The contrast (a), FWHM (b), and radial velocity values (c) are derived from the double-Gaussian best fits to the CCF_{PO} , and show similar values over the three nights. a–c, Visits 1, 2 and 3 are coloured in red, blue and orange, respectively. All error bars are 1σ .

Horizontal error bars correspond to the exposure time. Vertical dashed lines are the transit contacts. a, b, The width and contrast of the $\text{CCF}_{\text{DI}}^{\text{OT}}$ (horizontal dashed lines) are similar over the three visits. c, The dashed black line is the reloaded Rossiter–McLaughlin model corresponding to the best fit for the planet trajectory and the velocity field of the star.



Extended Data Figure 6 | Correlation diagram for the posterior probability distributions of the solid-body rotation model parameters. Green and blue lines show the two-dimensional confidence regions that contain 39.3% and 86.5% of the accepted steps, respectively.

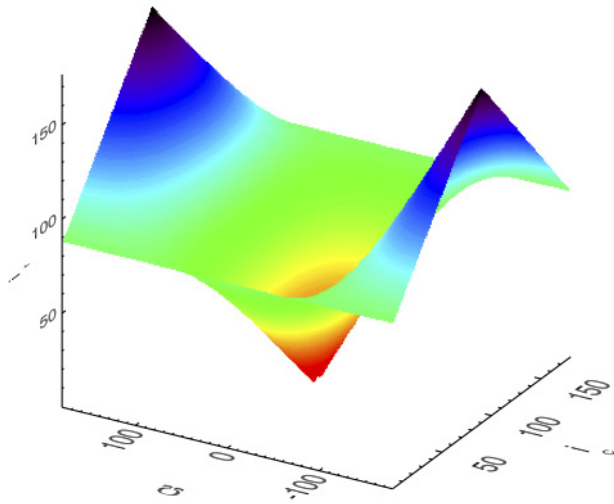
One-dimensional histograms correspond to the distribution projected on the space of each line parameter, with the orange dashed line limiting the 68.3% confidence interval. The red line and white point show median values.



Extended Data Figure 7 | Ground-based photometry of GJ 436.

a, Time series of GJ 436 nightly magnitude with transit points removed and normalized to the same seasonal mean. UTC, Coordinated Universal time; HJD, heliocentric Julian date. **b**, Frequency spectrum of the normalized observations with strongest peak at a photometric

period of 44.09 days, and secondary peaks corresponding to yearly aliases caused by the temporal sampling. **c**, Normalized data and best-fit sine curve (blue line) phased to $P_{\text{rot}} = 44.09$ days. The binned data (red squares) highlight the low-level brightness modulation of GJ 436 (peak-to-peak amplitude of 0.0032 mag).



Extended Data Figure 8 | Conditions on GJ 436b and GJ 436c orbital planes. For a given mutual inclination i_m (vertical axis), the acceptable properties for the orbital planes describe an oval ring in the (Ω, i_c) plane. Ω is the difference between the longitudes of the ascending nodes and i_c is the orbital inclination of GJ 436c.

Extended Data Table 1 | Log of GJ 436b transit observations

Visit number	1	2	3
Observation date	9 May 2007	18 March 2016	11 April 2016
Instrument	HARPS	HARPS-N	HARPS-N
Number of exposures	44	77	71
Exposures kept after color-correction	35	76	69
Before transit	6	63	20
During transit	11	9	9
After transit	18	4	40
Orbital phase of exposures that failed our color-correction	-0.018; -0.017; -0.009; -0.008; 0.022; 0.024; 0.025; 0.027; 0.028	0.016	0.083; 0.085
Orbital phase of exposures that failed our CCF detection criterion	-0.0065; 0.0050; 0.0065; 0.0079	-0.0079; 0.0070	-0.0079; 0.0069

Extended Data Table 2 | Properties of the GJ 436 system

Name	Fixed properties	Value	Reference
Stellar radius	R_*	$0.449 \pm 0.019 R_\odot$	Ref. 27
Stellar mass	M_*	$0.445 \pm 0.044 M_\odot$	Ref. 27
Effective temperature	T_{eff}	$3479 \pm 60 \text{ K}$	Ref. 27
Non-linear limb-darkening coefficients	u_1	1.47	Ref. 24
	u_2	-1.10	Ref. 24
	u_3	1.09	Ref. 24
	u_4	-0.42	Ref. 24
Semi-major axis	a_b/R_*	14.54 ± 0.14	Ref. 2
Mid-transit time	T_b	$2454865.084034 \pm 0.000035 \text{ BJD}$	Ref. 2
Orbital period	P_b	$2.64389803 \pm 2.6 \times 10^{-7} \text{ days}$	Ref. 2
Orbital inclination	i_b	$86.858 \pm 0.049 - 0.052^\circ$	Ref. 2
RV semi amplitude	K_b	$17.59 \pm 0.25 \text{ m s}^{-1}$	Ref. 2
Planet mass	M_b	$25.4 \pm 2.1 - 2.0 M_{\text{Earth}}$	Ref. 2
Transit depth	$(R_b/R_*)^2$	0.006819 ± 0.000028	Ref. 2
Eccentricity	e_b	0.1616 ± 0.004	Ref. 2
Argument of periastron	ω_b	$327.2 \pm 1.8 - 2.2^\circ$	Ref. 2
Name	Derived properties	Value	
Projected rotational velocity	$V_{\text{eq}} \sin i_*$	$0.330 \pm 0.091 - 0.066 \text{ km s}^{-1}$	
Projected obliquity	λ_b	$72 \pm 33 - 24^\circ$	
Stellar rotation period	P_{rot}	$44.09 \pm 0.08 \text{ days}$	
Stellar inclination	i_*	$39^{+13}_{-9}^\circ$	
		$141^{+9}_{-13}^\circ$	
3D obliquity	ψ_b	$80^{+21}_{-18}^\circ$	

Data are from refs 2, 24 and 27.

Enhancement and sign change of magnetic correlations in a driven quantum many-body system

Frederik Görg¹, Michael Messer¹, Kilian Sandholzer¹, Gregor Jotzu^{1,2}, Rémi Desbuquois¹ & Tilman Esslinger¹

Periodic driving can be used to control the properties of a many-body state coherently and to realize phases that are not accessible in static systems. For example, exposing materials to intense laser pulses makes it possible to induce metal–insulator transitions, to control magnetic order and to generate transient superconducting behaviour well above the static transition temperature^{1–6}. However, pinning down the mechanisms underlying these phenomena is often difficult because the response of a material to irradiation is governed by complex, many-body dynamics. For static systems, extensive calculations have been performed to explain phenomena such as high-temperature superconductivity⁷. Theoretical analyses of driven many-body Hamiltonians are more challenging, but approaches have now been developed, motivated by recent observations^{8–10}. Here we report an experimental quantum simulation in a periodically modulated hexagonal lattice and show that antiferromagnetic correlations in a fermionic many-body system can be reduced, enhanced or even switched to ferromagnetic correlations (sign reversal). We demonstrate that the description of the many-body system using an effective Floquet–Hamiltonian with a renormalized tunnelling energy remains valid in the high-frequency regime by comparing the results to measurements in an equivalent static lattice. For near-resonant driving, the enhancement and sign reversal of correlations is explained by a microscopic model of the system in which the particle tunnelling and magnetic exchange energies can be controlled independently. In combination with the observed sufficiently long lifetimes of the correlations in this system, periodic driving thus provides an alternative way of investigating unconventional pairing in strongly correlated systems experimentally^{7,9,10}.

The increasing demand for high-speed control of magnetic memory devices in the terahertz frequency regime has led to efforts to control the magnetic properties of materials optically, such as switching from antiferromagnetic to ferromagnetic ordering^{4,5}. To engineer suitable materials for future applications, it is desirable to gain a better understanding of the underlying microscopic processes. In this context, experiments using cold atoms provide an ideal platform for investigating driven many-body systems, owing to the slow timescales and the prospect of quantitative comparisons to theoretical predictions. So far, periodic modulation has been used in such set-ups to engineer effective Hamiltonians^{11,12}, which has enabled Hubbard parameters to be renormalized and classical magnetism to be studied in the high-frequency regime, as well as new features such as topological or spin-dependent band structures to be realized^{13–15}. By driving interacting systems^{16,17}, charge and spin degrees of freedom can both be influenced by addressing density-dependent processes individually^{18–20}. Until now, the measurement of magnetic correlations in driven optical lattices has remained an open challenge. An experimental difficulty lies in the heating associated with the periodic modulation of a many-body system, which can destroy correlations, especially in the near-resonant regime^{14,21,22}.

We perform our experiments using a degenerate Fermi gas consisting of $3.0(2) \times 10^4$ (10% systematic error) ultracold ⁴⁰K atoms prepared in a balanced mixture of two internal states, denoted as \uparrow and \downarrow

(see Methods). The atoms are loaded into an optical superlattice with a tunable geometry and anisotropic tunnelling rates, whereby the horizontal links in the x direction (t_x) are stronger than those in the y and z directions (t_{yz} ; Fig. 1c). In the x – z plane, the lattice consists of hexagonal layers, which are stacked in the y direction. We modulate the lattice position in the x direction periodically in time with a displacement amplitude A at a frequency of $\omega/(2\pi)$, which is achieved by moving the retroreflecting mirror of the optical lattice using a piezoelectric actuator (Fig. 1a).

Our system is well described by the driven Fermi–Hubbard model:

$$\hat{H}(\tau) = - \sum_{\langle i,j \rangle, \sigma} t_{ij} \hat{c}_{i\sigma}^\dagger \hat{c}_{j\sigma} + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} + \sum_{i,\sigma} [f_i(\tau) + V_i] \hat{n}_{i\sigma} \quad (1)$$

where $\hat{c}_{i\sigma}^\dagger$, $\hat{c}_{i\sigma}$ and $\hat{n}_{i\sigma}$ are the fermionic creation, annihilation and number operators, respectively, at site $i = (i_x, i_y, i_z)$ in spin state $\sigma \in \{\uparrow, \downarrow\}$.

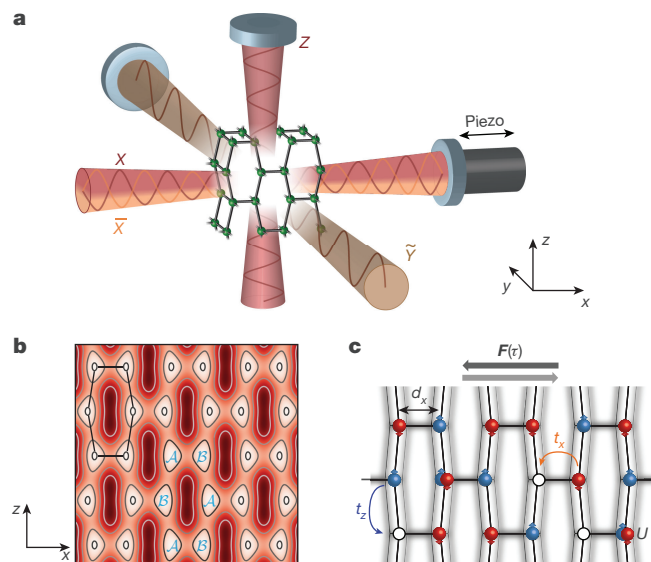


Figure 1 | Experimental set-up. **a**, Optical set-up used to create the three-dimensional lattice geometry. The beams X and Z are interfering, whereas \bar{X} and \bar{Y} are frequency-detuned. A piezoelectric actuator sinusoidally modulates the position of the retroreflecting mirror in the x direction.

b, Lattice potential (colour scale, lighter red corresponds to a lower potential depth) in the x – z plane. The lattice consists of A and B sublattices, and a hexagonal unit cell is superimposed. **c**, Tight-binding representation of the lattice potential in the x – z plane. The system is described by a driven Fermi–Hubbard model, with anisotropic tunnelling energies $t_x > t_z$, owing to the shorter length d_x of the horizontal bonds.

Atoms in different spin states (red and blue, arrows) interact via an on-site interaction U . In a co-moving frame, the modulation of the lattice position (indicated by grey lattices in the background) corresponds to a linear force $F(\tau)$ in the x direction with an amplitude of $\hbar\omega K_0/d_x$, which primarily influences the horizontal bonds ($F(\tau) = (\hbar\omega K_0/d_x)\cos(\omega\tau)\mathbf{e}_x$; Methods).

¹Institute for Quantum Electronics, ETH Zurich, 8093 Zurich, Switzerland. ²Max Planck Institute for the Structure and Dynamics of Matter, 22761 Hamburg, Germany.

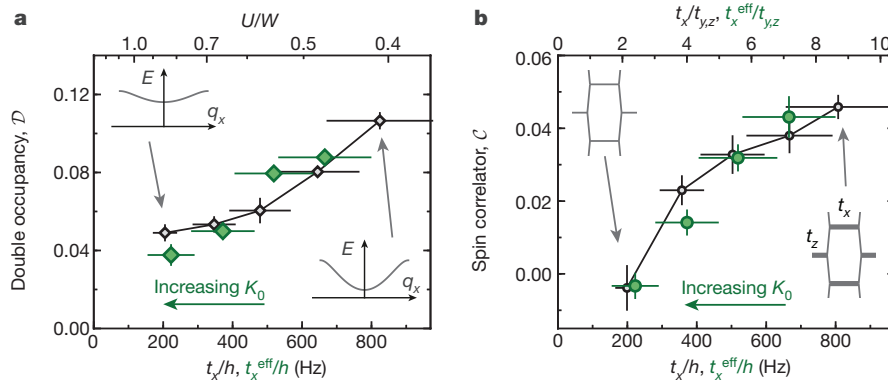


Figure 2 | Description of the driven system by an effective Hamiltonian in the high-frequency regime. **a**, Double occupancy \mathcal{D} as a function of the effective horizontal tunnelling energy $t_x^{\text{eff}}(K_0) = t_x \mathcal{J}_0(K_0)$ for a driven system (green), and results obtained from an experimental quantum simulation in a static configuration with horizontal tunnelling t_x (black). The insets show cuts through the non-interacting band structure (E , energy) as a function of the quasi-momentum in the x direction q_x . The reduction in the bandwidth W leads to a lower double occupancy, indicating the crossover to a Mott-insulating state. **b**, Spin–spin correlations \mathcal{C} as a function of the (effective)

horizontal tunnelling energy for the driven case (green) and an equivalent static configuration (black). The renormalization of the tunnelling energy leads to a reduction in lattice anisotropy $t_x^{\text{eff}}/t_{y,z}$ (see insets), which reduces the magnetic correlations on the horizontal link. The transverse tunnelling energies are $t_y/\hbar = 125(8)$ Hz and $t_z/\hbar = 78(8)$ Hz and the interaction is set to $U/\hbar = 0.93(2)$ kHz. Horizontal error bars reflect the uncertainty in the lattice depth; data points and vertical error bars in **a** (**b**) denote the mean and standard error of 4 (10) individual measurements at different times within one driving period (see Methods).

Here, t_{ij} denotes the tunnelling rate between nearest neighbours $\langle i, j \rangle$, U the repulsive on-site interaction and V_i an overall harmonic trapping potential. The time-dependent force is expressed as $f_i(\tau) = m\omega^2 x_i \cos(\omega\tau)$, where m is the mass of the atoms and $x_i = \langle \hat{x} \rangle_i$ is the x position of the Wannier function on site i . Therefore, the driving is used primarily to address the bonds in the x direction (Methods). To characterize the many-body state in the lattice, we measure the fraction of atoms on doubly occupied sites

$$\mathcal{D} = \frac{2}{N} \sum_{i \in \mathcal{A}, \mathcal{B}} \langle \hat{n}_{i1} \hat{n}_{i2} \rangle$$

and the nearest-neighbour spin–spin correlator

$$\mathcal{C} = -\frac{1}{N} \sum_{i \in \mathcal{A}} (\langle \hat{S}_i^x \hat{S}_{i+e_x}^x \rangle + \langle \hat{S}_i^y \hat{S}_{i+e_x}^y \rangle)$$

on the horizontal links along the x direction. (Here N is the total number of atoms, e_x is the unit vector in the x direction, which connects the sites of the \mathcal{A} and \mathcal{B} sublattices, and \hat{S}_i represents the standard spin vector operator on site i .) The observables are averaged spatially over the inhomogeneous density distribution in the harmonic trap, which has a geometric mean trapping frequency of $\bar{\omega}_{\text{trap}}/(2\pi) = 84(2)$ Hz, and over one oscillation cycle of the periodic modulation, as indicated by $\langle \dots \rangle$ (see Methods).

In a first experiment, we investigate the regime in which the driving frequency is much higher than all microscopic energy scales of the system, that is, the tunnelling t and interaction energy U ($\hbar\omega \gg t, U$). In the non-interacting case, the modulation renormalizes the horizontal tunnelling rate by a zeroth-order Bessel function (\mathcal{J}_0) and the system can be described by an effective tunnelling energy

$$t_x^{\text{eff}}(K_0) = t_x \mathcal{J}_0(K_0) \quad (2)$$

where $K_0 = m\omega d_x/\hbar$ is the normalized driving amplitude, with d_x the length of the horizontal bonds (Fig. 1c)¹⁵. However, it is not clear *a priori* whether this simple description remains accurate in the many-body context¹². To verify this, we compare our measurements in the driven system to results obtained using an experimental quantum simulation in a static lattice with a variable tunnelling rate t_x . The reliability of our experiment as a quantum simulator for the magnetic properties of the Hubbard model has previously been benchmarked through quantitative comparisons with state-of-the-art numerical calculations^{23,24}. To enter

the driven regime in the experiment, we ramp up the lattice modulation amplitude linearly to a final value K_0 within 2 ms, at a frequency of $\omega/(2\pi) = 6$ kHz. Afterwards, we allow for an additional equilibration time of 5 ms before the measurement, during which we maintain a fixed modulation amplitude.

The resulting double occupancies and spin correlations agree well for the driven and static cases, as shown in Fig. 2. This supports the validity of the description of the many-body system by an effective Hamiltonian with a tunnelling rate $t_x^{\text{eff}}(K_0)$. For lower tunnelling energies, the double occupancy decreases as a result of the reduction in the bandwidth W . Therefore, for increasing driving amplitude, the system enters the Mott regime¹⁶. The modulation not only changes the bandwidth, but also the anisotropy of the lattice, because the ratio $t_x^{\text{eff}}(K_0)/t_{y,z}$ decreases for increasing driving amplitude. This effect manifests in the spin correlator on the horizontal link, which decreases for a weaker anisotropy of the underlying lattice, as observed in previous measurements²⁵. When driving for longer times, we find that the lifetime of correlations is reduced to 14(5) ms at $K_0 = 1.26(4)$, compared to 92(16) ms in the static case. Nevertheless, this allows us to observe comparable levels of correlations in the driven and static cases on experimental timescales.

Whereas an off-resonant modulation scheme typically leads to a renormalization of pre-existing parameters, physics that is not accessible in static systems arises for a near-resonant drive. For example, extended terms such as density-dependent tunnelling energies can be engineered, which are not present in the single-band Hubbard model^{18–20}. To investigate this regime, we set a large on-site interaction close to the driving frequency ($U \approx \hbar\omega$, $l \in \mathbb{Z}$) and ramp up the modulation at a frequency of either 3 kHz or 6 kHz within 3.3 ms or 2 ms, respectively. We observe that the effective states in the driven Hamiltonian contain a higher fraction of double occupancies if $U \approx \hbar\omega$ (Fig. 3a).

Strikingly, we find that the magnetic correlations on the horizontal links depend on both the sign and magnitude of the modulation detuning $\delta = \hbar\omega - U$ (Fig. 3b). For a red-detuned drive ($\delta < 0$), correlations are increased compared to the static case if $|\delta|$ is of the order of a few tunnelling energies t_x . By contrast, when choosing $\delta > 0$, the sign of the spin–spin correlator inverts; that is, the system exhibits ferromagnetic correlations on neighbouring sites in the x direction. If we set a fixed interaction strength and vary the amplitude of the modulation, then we find that correlations increase for $\delta < 0$ and $K_0 \approx 1.3$, before they eventually decrease again (Fig. 3c). For $\delta > 0$, a critical value of the driving strength is required for the system to develop ferromagnetic

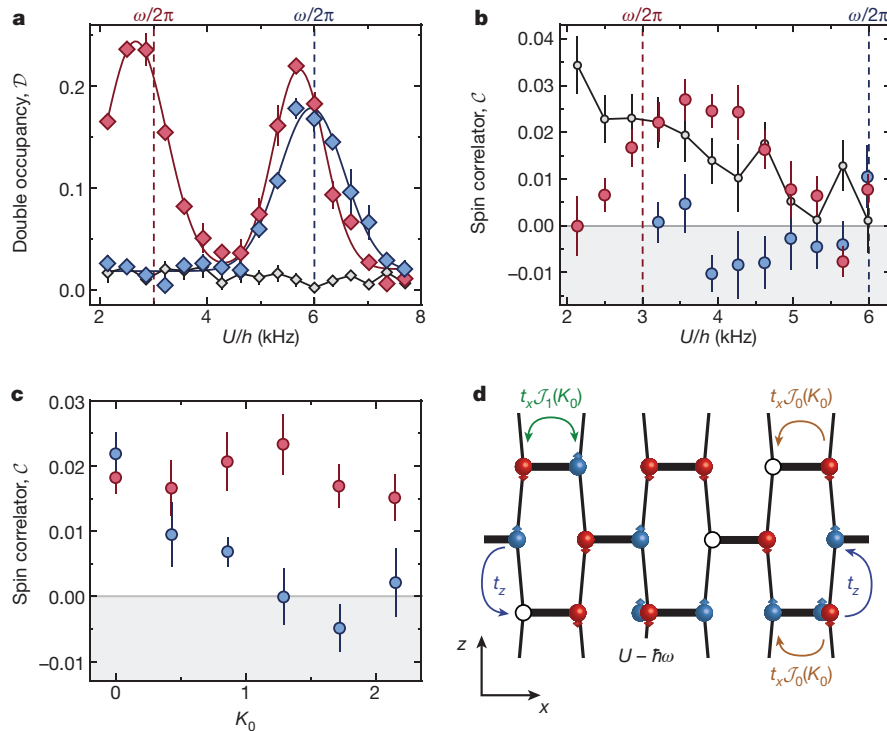


Figure 3 | Enhancement and sign reversal of magnetic correlations by near-resonant driving. **a**, Double occupancy as a function of on-site interaction U for the static case (black) and for driving frequencies of $\omega/(2\pi) = 3$ kHz (red) or 6 kHz (blue) with a modulation amplitude of $K_0 = 1.30(3)$. Around the resonances (vertical dashed lines), the effective states in the driven Hamiltonian contain a higher number of double occupancies. Solid lines are (double) Gaussian fits to the data. **b**, Spin–spin correlations on the horizontal link as a function of U for the same parameters as in **a**. For $U > \hbar\omega$ (red), antiferromagnetic correlations are enhanced compared to the static case (black) for a broad range of interactions. When $U < \hbar\omega$ (blue), the correlator changes sign and the system develops ferromagnetic correlations. **c**, Spin–spin correlations as a function of driving amplitude K_0 for $\omega/(2\pi) = 3$ kHz

and $U/h = 3.8(1)$ kHz (red) or $\omega/(2\pi) = 6$ kHz and $U/h = 4.4(1)$ kHz (blue). For $U > \hbar\omega$, antiferromagnetic correlations increase around $K_0 \approx 1.3$. For $\hbar\omega > U$, correlations become ferromagnetic beyond a critical modulation amplitude. The tunnelling rates are set to $(t_x, t_y, t_z)/h = (570(110), 125(8), 85(8))$ Hz. Data points and error bars in **a** (**b** and **c**) denote the mean and standard error of 4 (10) individual measurements at different times within one driving period (see Methods). **d**, In the near-resonant case ($U \approx \hbar\omega$), the driven system can be described by an effective Hamiltonian in which tunnelling processes that do not change the number of double occupancies are renormalized by $\mathcal{J}_0(K_0)$ (brown). By contrast, the creation of doublon–holon pairs is resonantly enhanced and is determined by the first-order Bessel function $\mathcal{J}_1(K_0)$ (green). The effective interaction of the system becomes $U - \hbar\omega$.

correlations. We also study the time dependence of the magnetic properties, by varying the modulation time after the ramp up of the drive. We find that it takes a few milliseconds for correlations to increase or change sign, but that they ultimately approach zero when driving for long times as a result to heating of the cloud (Extended Data Fig. 1). The lifetime of magnetic correlations as extracted from an exponential fit to the long-time behaviour changes from 82(34) ms in the static case to 12(4) ms at $K_0 = 1.30(3)$. In addition, we observe the fast dynamics within one period of the drive (the so-called micromotion) in our measurement regime (Extended Data Fig. 2). Finally, we investigate the adiabaticity of the preparation protocol by reverting the driving ramp and find that correlations return only partially to their static values (Extended Data Fig. 3).

To obtain an understanding of the observed phenomena at the microscopic scale, we perform a Floquet analysis on the time-periodic Hamiltonian in equation (1) in the near-resonant driving regime with $t \ll U \approx \hbar\omega$. For that, we switch to a rotating frame with respect to the operator

$$\hat{R}(\tau) = \exp \left\{ i \sum_j \left[l\omega \hat{n}_{j\uparrow} \hat{n}_{j\downarrow} + \sum_{\sigma} F_j(\tau) \hat{n}_{j\sigma} \right] \right\}$$

where

$$F_j(\tau) = \frac{1}{\hbar} \int_0^{\tau} f_j(\tau') d\tau'$$

In this frame, the tunnelling on the horizontal bonds is to lowest order in $1/\omega$ described by the effective Hamiltonian

$$\hat{H}_{t_x}^{\text{eff}} = -t_x \sum_{\substack{i \in A, \sigma \\ j = i + e_x}} \left[\mathcal{J}_0(K_0) \hat{a}_{i\sigma} \hat{b}_{j\sigma}^{\dagger} + \mathcal{J}_1(K_0) \hat{b}_{i\sigma}^{\dagger} \hat{c}_{j\sigma} \right] + \text{h.c.} \quad (3)$$

where $\uparrow = \downarrow$ and vice versa^{26–28}. Here, the effective tunnelling energy is density-dependent: hopping processes that do not change the number of double occupancies as described by the operator $\hat{a}_{i\sigma} = (1 - \hat{n}_{i\sigma})(1 - \hat{n}_{j\sigma}) + \hat{n}_{i\sigma} \hat{n}_{j\sigma}$ are renormalized by $\mathcal{J}_0(K_0)$. In contrast, the creation or annihilation of doublon–holon pairs corresponding to $\hat{b}_{i\sigma}^{\dagger} = (-1)^i (1 - \hat{n}_{i\sigma}) \hat{n}_{j\sigma} + \hat{n}_{i\sigma} (1 - \hat{n}_{j\sigma})$ become resonantly restored with an amplitude $t_x \mathcal{J}_1(K_0)$ (Fig. 3d). In addition, the effective interaction $U^{\text{eff}} = U - \hbar\omega = -\delta_i$ is given by the detuning from the l -photon resonance δ_i . In this picture, we can understand the creation of double occupancies for small δ_i shown in Fig. 3a as the system becoming effectively more weakly interacting.

The magnetic properties of the many-body state are altered substantially in the effective Hamiltonian in equation (3) because at the microscopic scale the superexchange process that leads to spin–spin interactions involves two virtual hopping processes determined by $\mathcal{J}_l(K_0)$, in which a double occupancy at energy U^{eff} is created and annihilated. Therefore, the exchange energy J_{ex} , which is the energy splitting between a spin singlet and triplet state on the horizontal bonds, will depend on both the modulation amplitude K_0 and the detuning δ . It can even change sign for $\delta > 0$, because in this case the effective

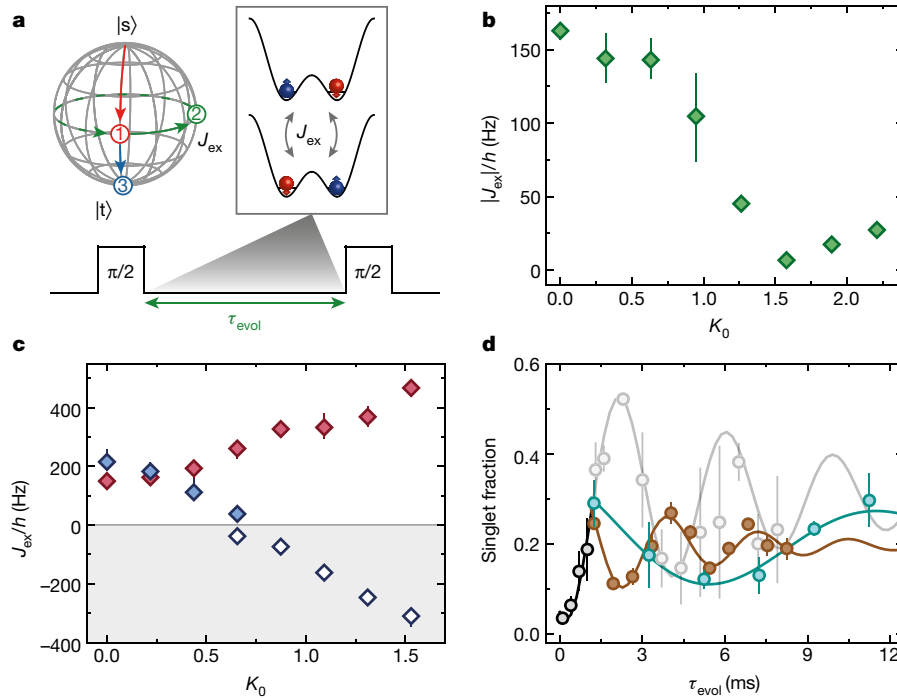


Figure 4 | Magnetic exchange energy for off- and near-resonant driving.

a, The exchange J_{ex} is measured by preparing local singlet states $|s\rangle$ on isolated double wells. In a Ramsey-type sequence, a superposition between the singlet $|s\rangle$ and triplet $|t\rangle$ states is first created by performing a $\pi/2$ pulse (red arrow) with a magnetic field gradient. The exchange oscillation (green arrow; the solid component represents the one-quarter-oscillation evolution time used in **d**) is then triggered by suddenly lowering the barrier in the double well. Finally, after a variable evolution time τ_{evol} , a second $\pi/2$ pulse (blue arrow) is applied and the final singlet fraction is measured, which oscillates at a frequency $|J_{ex}|$. **b**, Magnetic exchange in the off-resonant driving regime for $\omega/(2\pi) = 8$ kHz, $t_x/h = 350(50)$ Hz and $U/h = 2.1(1)$ kHz as a function of driving amplitude. J_{ex} decreases with K_0 as expected for a renormalized tunnelling rate t_x^{eff} . **c**, Exchange energy for near-resonant modulation with $\omega/(2\pi) = 8$ kHz, $t_x/h = 640(90)$ Hz and $U/h = 9.1(1)$ kHz (red) or $U/h = 6.5(1)$ kHz (blue) as a function of K_0 . Red-

interaction becomes attractive^{8–10,29} (Extended Data Fig. 4, Methods). We measure J_{ex} between neighbouring sites directly in the experiment using our optical lattice with tunable geometry. For that, we disconnect individual pairs of sites in the x direction from each other by raising the potential barrier in the y and z directions, so that the coupling $t_{yz}/h < 2$ Hz becomes negligible, and measure the exchange energy in a Ramsey-type sequence (Fig. 4a)^{29,30}.

The results of the measurements in the off- and near-resonant driving regimes for a modulation frequency of $\omega/(2\pi) = 8$ kHz are shown in Fig. 4. In the case of high-frequency modulation with $t_x \ll U \ll \hbar\omega$, the tunnelling is renormalized according to equation (2) and the exchange energy decreases as a function of the driving amplitude as $J_{ex} \approx 4t_x^2 J_0^2(K_0)/U$ (Fig. 4b). By contrast, in the near-resonant regime, the system is to lowest order described by the tunnelling process in equation (3) and we observe an increasing exchange energy as a function of the modulation strength for $\delta < 0$ (Fig. 4c). At $K_0 \approx 1.6$ it reaches a level about three times higher than in the static case. If $\delta > 0$, J_{ex} vanishes at a critical modulation amplitude of $K_0 \approx 0.7$ and changes sign for stronger driving. To demonstrate that the exchange becomes negative for large K_0 , we first perform a quarter oscillation in the static double well, followed by a sudden switch on of the modulation with $K_0 > 0.7$ (ref. 29). Because the exchange in the driven double well is ferromagnetic, it inverts its rotation direction on the Bloch sphere, which leads to an oscillation phase shifted by π compared to the static case (Fig. 4d).

The dependence of the exchange energy on the driving frequency and strength provides a microscopic explanation for the phenomena

detuned driving ($U > \hbar\omega$) enhances the magnetic exchange for increasing driving amplitude. For $U < \hbar\omega$, J_{ex} vanishes at a critical value $K_0 \approx 0.7$ and becomes negative for stronger driving (open symbols). The sign of the exchange is measured as shown in **d**. For $K_0 \approx 0.7$, the oscillation is too slow to determine the sign of J_{ex} . Mean values in **b** and **c** are derived from a sinusoidal fit to the data; errors denote the standard deviation obtained from a resampling method (see Methods). **d**, Sign change of the exchange energy for $U < \hbar\omega$, as indicated by the singlet fraction. The singlet fraction is shown as a function of evolution time, with $U/h = 6.5(1)$ kHz in the static case (black, grey) or after a sudden switch on of the modulation with $K_0 = 0.88(1)$ (cyan) or $K_0 = 1.31(2)$ (brown) after a quarter exchange oscillation, and with the other parameters as in **c**. Owing to the sign reversal of J_{ex} , the rotation direction on the Bloch sphere is reversed. Solid lines are damped sinusoidal fits to the data. Error bars denote the standard deviation of 3 measurements.

observed in the many-body system. In the off-resonant case, the magnetic exchange decreases with increasing modulation amplitude, which reduces the lattice anisotropy and therefore the correlations on the x bonds (Fig. 2b). If the interaction energy U comes close to, but is still lower than the driving frequency, then resonant effects start to dominate and the magnetic exchange inverts its sign, leading to ferromagnetic correlations in the many-body system as observed in Fig. 3b, c. For $U \gtrsim \hbar\omega$, the exchange energy increases with K_0 , which can enhance antiferromagnetic correlations for several reasons. First, the anisotropy is increased because the ratio $J_{ex}^x/J_{ex}^{y,z}$ becomes larger, which makes it more favourable to redistribute entropy onto the weak links in the y and z directions^{24,25}. Second, while the exchange is increased, the single-particle tunnelling energy is renormalized as $t_{x,single} = t_x J_0(K_0)$ in the effective Hamiltonian; see equation (3). Owing to the isolated nature of the entire system, the reduction of $t_{x,single}$ leads to an entropy redistribution in the trap and lowers the absolute temperature, which enhances magnetic correlations globally. Last, when the ratio $J_{ex}/t_{x,single}$ increases, it becomes more favourable for two atoms to pair and form a singlet state in the low filled regions of the trap instead of delocalizing far apart⁹. This process plays an important part in the context of high-temperature superconductivity, and the independent control of the exchange and tunnelling energies opens up the possibility of investigating d -wave pairing in the t - J model⁷. Further theoretical studies will be necessary to determine the degree to which these three effects are responsible for the enhancement of antiferromagnetic correlations in the many-body system.

Having shown that near-resonant driving can be used to increase or reverse the sign of magnetic correlations, the low-energy scales in systems of cold atoms enable further investigations of the timescales involved and the possible existence of pre-thermalized states in future experiments²¹. Remarkably, the lifetime of correlations in the driven many-body system was found to be sufficiently long that they could be observed even in the near-resonant driving regime. To investigate this further, the entropy increase could be studied systematically as a function of the energy scales involved and the connectivity of the underlying lattice geometry. Furthermore, by additionally imprinting complex phases on the density-assisted tunnelling energies, dynamical gauge fields and anyonic statistics could be engineered²⁶.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 July; accepted 7 November 2017.

1. Kirilyuk, A., Kimel, A. V. & Rasing, T. Ultrafast optical manipulation of magnetic order. *Rev. Mod. Phys.* **82**, 2731–2784 (2010).
2. Nicoletti, D. & Cavalleri, A. Nonlinear light-matter interaction at terahertz frequencies. *Adv. Opt. Photonics* **8**, 401–464 (2016).
3. Rini, M. *et al.* Control of the electronic phase of a manganite by mode-selective vibrational excitation. *Nature* **449**, 72–74 (2007).
4. Mariager, S. O. *et al.* Structural and magnetic dynamics of a laser induced phase transition in FeRh. *Phys. Rev. Lett.* **108**, 087201 (2012).
5. Li, T. *et al.* Femtosecond switching of magnetism via strongly correlated spin-charge quantum excitations. *Nature* **496**, 69–73 (2013).
6. Mitrano, M. *et al.* Possible light-induced superconductivity in K_3C_{60} at high temperature. *Nature* **530**, 461–464 (2016).
7. Dagotto, E. Correlated electrons in high-temperature superconductors. *Rev. Mod. Phys.* **66**, 763–840 (1994).
8. Mentink, J. H., Balzer, K. & Eckstein, M. Ultrafast and reversible control of the exchange interaction in Mott insulators. *Nat. Commun.* **6**, 6708 (2015).
9. Coulthard, J., Clark, S. R., Al-Assam, S., Cavalleri, A. & Jaksch, D. Enhancement of superexchange pairing in the periodically driven Hubbard model. *Phys. Rev. B* **96**, 085104 (2017).
10. Kitamura, S. & Aoki, H. η -pairing superfluid in periodically-driven fermionic Hubbard model with strong attraction. *Phys. Rev. B* **94**, 174503 (2016).
11. Goldman, N. & Dalibard, J. Periodically driven quantum systems: effective Hamiltonians and engineered gauge fields. *Phys. Rev. X* **4**, 031027 (2014).
12. Bukov, M., D'Alessio, L. & Polkovnikov, A. Universal high-frequency behavior of periodically driven systems: from dynamical stabilization to Floquet engineering. *Adv. Phys.* **64**, 139–226 (2015).
13. Struck, J. *et al.* Quantum simulation of frustrated classical magnetism in triangular optical lattices. *Science* **333**, 996–999 (2011).
14. Jotzu, G. *et al.* Creating state-dependent lattices for ultracold fermions by magnetic gradient modulation. *Phys. Rev. Lett.* **115**, 073002 (2015).
15. Eckardt, A. Colloquium: Atomic quantum gases in periodically driven optical lattices. *Rev. Mod. Phys.* **89**, 011004 (2017).
16. Zenesini, A., Lignier, H., Ciampini, D., Morsch, O. & Arimondo, E. Coherent control of dressed matter waves. *Phys. Rev. Lett.* **102**, 100403 (2009).
17. Parker, C. V., Ha, L.-C. & Chin, C. Direct observation of effective ferromagnetic domains of cold atoms in a shaken optical lattice. *Nat. Phys.* **9**, 769–774 (2013).
18. Ma, R. *et al.* Photon-assisted tunneling in a biased strongly correlated Bose gas. *Phys. Rev. Lett.* **107**, 095301 (2011).
19. Meinert, F., Mark, M. J., Lauber, K., Daley, A. J. & Nägerl, H.-C. Floquet engineering of correlated tunneling in the Bose-Hubbard model with ultracold atoms. *Phys. Rev. Lett.* **116**, 205301 (2016).
20. Desbuquois, R. *et al.* Controlling the Floquet state population and observing micromotion in a periodically driven two-body quantum system. *Phys. Rev. A* **96**, 053602 (2017).
21. Kuwahara, T., Mori, T. & Saito, K. Floquet-Magnus theory and generic transient dynamics in periodically driven many-body quantum systems. *Ann. Phys.* **367**, 96–124 (2016).
22. Abanin, D. A., De Roeck, W., Ho, W. W. & Huvneers, F. Effective Hamiltonians, prethermalization, and slow energy absorption in periodically driven many-body systems. *Phys. Rev. B* **95**, 014112 (2017).
23. Sciolia, B. *et al.* Competition of spin and charge excitations in the one-dimensional Hubbard model. *Phys. Rev. A* **88**, 063629 (2013).
24. Imriška, J. *et al.* Thermodynamics and magnetic properties of the anisotropic 3D Hubbard model. *Phys. Rev. Lett.* **112**, 115301 (2014).
25. Greif, D., Jotzu, G., Messer, M., Desbuquois, R. & Esslinger, T. Formation and dynamics of antiferromagnetic correlations in tunable optical lattices. *Phys. Rev. Lett.* **115**, 260401 (2015).
26. Bermudez, A. & Porras, D. Interaction-dependent photon-assisted tunneling in optical lattices: a quantum simulator of strongly-correlated electrons and dynamical gauge fields. *New J. Phys.* **17**, 103021 (2015).
27. Itin, A. & Katsnelson, M. Effective Hamiltonians for rapidly driven many-body lattice systems: induced exchange interactions and density-dependent hoppings. *Phys. Rev. Lett.* **115**, 075301 (2015).
28. Bukov, M., Kolodrubetz, M. & Polkovnikov, A. Schrieffer-Wolff transformation for periodically driven systems: strongly correlated systems with artificial gauge fields. *Phys. Rev. Lett.* **116**, 125301 (2016).
29. Trotzky, S. *et al.* Time-resolved observation and control of superexchange interactions with ultracold atoms in optical lattices. *Science* **319**, 295–299 (2008).
30. Chen, Y.-A. *et al.* Controlling correlated tunneling and superexchange interactions with ac-driven optical lattices. *Phys. Rev. Lett.* **107**, 210405 (2011).

Acknowledgements We thank D. Abanin, D. Greif, D. Jaksch, M. Landini, Y. Murakami, N. Tsuji, P. Werner and W. Zwerger for discussions. We acknowledge SNF (Project Number 200020_169320 and NCCR-QSIT), Swiss State Secretary for Education, Research and Innovation Contract No. 15.0019 (QUIC) and ERC advanced grant TransQ (Project Number 742579) for funding.

Author Contributions All authors contributed extensively to the work presented in this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to T.E. (esslinger@phys.ethz.ch).

Reviewer Information Nature thanks J. Freericks and D. Huse for their contribution to the peer review of this work.

METHODS

Optical lattice. The tunable three-dimensional optical lattice is created by a combination of four orthogonal, retroreflected laser beams of wavelength $\lambda = 1,064$ nm, as shown in Fig. 1a. Whereas the \bar{X} and \bar{Y} beams are interfering and actively phase-locked to $\varphi = 0.00(3)\pi$, the X and Z beams are non-interfering, owing to a frequency detuning. Our optical set-up is described by the following potential³¹:

$$\begin{aligned} V(x, y, z) = & -V_{\bar{X}}\cos^2(kx + \theta/2) - V_X\cos^2(kx) \\ & -V_{\bar{Y}}\cos^2(ky) - V_Z\cos^2(kz) \\ & -2\alpha\sqrt{V_{\bar{X}}V_Z}\cos(kx)\cos(kz)\cos(\varphi) \end{aligned} \quad (4)$$

with $k = 2\pi/\lambda$ and $V_{\bar{X},X,\bar{Y},Z}$ the lattice depths in units of the recoil energy $E_R = \hbar^2/(2m\lambda^2)$ of each laser beam in the three directions x , y and z (\hbar is the Planck constant and m the mass of the atoms). The lattice potential is adjusted to fix $\theta = 1.000(2)\pi$. We calibrate the visibility of the interference term $\alpha = 0.92(1)$ with amplitude modulation of the lattice depth for different configurations of the optical potential using a ⁸⁷Rb Bose–Einstein condensate. To calibrate the individual lattice depths $V_{\bar{X},X,\bar{Y},Z}$ we perform Raman–Nath diffraction on the Bose–Einstein condensate. For the calculation of tight-binding parameters, we include a systematic error of 3% for all lattice depths.

Preparation of the degenerate Fermi gas in the optical lattice. The starting point of our experiment is a balanced mixture of the $F=9/2$, $m_F = -9/2$ and $F=9/2$, $m_F = -7/2$ hyperfine states of ⁴⁰K, confined in an optical harmonic trap. We evaporatively cool the mixture to a quantum degenerate cloud with a repulsive s -wave scattering length of $115.6(8)a_0$ (a_0 denotes the Bohr radius). After the evaporation, we end up with about $3.0(2) \times 10^4$ (10% systematic error) atoms at a temperature of $T/T_F = 0.07(1)$ (T_F denotes the Fermi temperature, see Extended Data Table 1 for details). Afterwards, we either keep a mixture of the $F=9/2$, $m_F = -9/2$ and $F=9/2$, $m_F = -7/2$ hyperfine states to access attractive or weak repulsive interactions with scattering lengths $-3,000a_0 < a < 150a_0$ (measurements in Figs 2 and 4b and for the initial preparation of isolated double wells in Fig. 4), or transfer the $F=9/2$, $m_F = -7/2$ state to the $F=9/2$, $m_F = -5/2$ state with a radio-frequency sweep to access large repulsive scattering lengths above $200a_0$ (measurements in Figs 3 and 4c, d). For this mixture, we obtain temperatures of $T/T_F = 0.12(2)$ in the harmonic trap. The interactions can be tuned via two magnetic Feshbach resonances located at a field of 202.1 G (for $m_F = -9/2$ and $m_F = -7/2$) or 224.2 G (for $m_F = -9/2$ and $m_F = -5/2$). From this point, two distinct schemes are used to prepare atoms either in a three-dimensional hexagonal lattice (Figs 2, 3) or in isolated double wells (Fig. 4). To load a many-body state into the hexagonal lattice, we first ramp up the power of all lattice beams in 50 ms to an intermediate value. In this configuration, the tunnelling energies are close to the final configuration with $(t_x, t_y, t_z)/\hbar = (550(30), 143(8), 156(9))$ Hz, but the horizontal link across the hexagonal unit cell still has a finite value of $70(3)$ Hz. In addition, the mean trap frequency is only $\bar{\omega}_{\text{trap}} = 68(2)$ Hz. In the second step, we ramp up the power in all beams in 20 ms to the final configuration (Extended Data Table 1). To load isolated double wells, we first tune the interactions to a large attractive value of $-3,000(600)a_0$; see ref. 20 for more details. In short, the atoms are first loaded into the lowest band of a checkerboard configuration with $V_{\bar{X},X,\bar{Y},Z} = [0, 3, 7, 3]E_R$ using an S-shaped lattice ramp of 200 ms. Owing to the large attractive interactions during the loading process, 68(3)% of the atoms form double occupancies. In the second step, we tune the scattering length to $115.6(8)a_0$ and split each lattice site by linearly increasing $V_{\bar{X}}$ and decreasing V_X to a $V_{\bar{X},X,\bar{Y},Z} = [30, 0, 30, 30]E_R$ cubic configuration within 10 ms. During the splitting process, the double occupancies in the checkerboard lattice are transformed into singlet states $|s\rangle = (|\uparrow, \downarrow\rangle - |\downarrow, \uparrow\rangle)/\sqrt{2}$ in the cubic lattice.

Detection methods. The detection scheme of double occupancies and nearest-neighbour spin–spin correlations follows closely the procedure used in previous work^{25,32}. To characterize the atomic state, we first freeze the evolution by quenching the lattice to $V_{\bar{X},X,\bar{Y},Z} = [30, 0, 30, 30]E_R$ within 100 μ s. To detect double occupancies, we ramp the magnetic field close to the magnetic Feshbach resonance of the $m_F = -9/2$ and $m_F = -7/2$ mixture. We then selectively transfer one of the atoms sitting on doubly occupied sites from the $m_F = -7/2$ state to the $m_F = -5/2$ state, or vice versa, via a radio-frequency sweep by using the interaction shift. The number of atoms in the different Zeeman sublevels can then be determined by applying a Stern–Gerlach pulse during the time-of-flight imaging. For the measurement of spin–spin correlations, we apply a magnetic-field gradient after the lattice freeze. This leads to coherent oscillations between the magnetic singlet state $|s\rangle = (|\uparrow, \downarrow\rangle - |\downarrow, \uparrow\rangle)/\sqrt{2}$ and triplet state $|t\rangle = (|\uparrow, \downarrow\rangle + |\downarrow, \uparrow\rangle)/\sqrt{2}$ on neighbouring sites in the x direction. The singlet fraction p_s can be determined by merging adjacent lattice sites by going to a $V_{\bar{X},X,\bar{Y},Z} = [0, 30, 30, 30]E_R$ checkerboard configuration within 10 ms. This procedure transforms the singlet into a double

occupancy in the single well, which can again be measured as outlined above. The triplet fraction p_t is obtained by applying a π pulse with the magnetic-field gradient and subsequently measuring the singlet fraction. The spin–spin correlation is then obtained as $C = -\langle \hat{S}_i^x \hat{S}_{i+1}^x \rangle - \langle \hat{S}_i^y \hat{S}_{i+1}^y \rangle = (p_s - p_t)/2$. We average all observables over one period $T = 2\pi/\omega$ of the drive to be insensitive to the micro-motion. For that, we vary slightly the total duration of the modulation between different measurements by multiples of $T/4$ to sample different phases of the modulation cycle. For the measurement of double occupancies in the hexagonal lattice (Figs 2a, 3a) we sample four different times during the modulation cycle, whereas for the magnetic correlations (Figs 2b, 3b, c, Extended Data Figs 1, 2) we measure for five different times and take each data point two or three times (see captions for the exact number of measurements). For the measurements performed in the isolated double wells (Fig. 4) the observables were not averaged over one driving period because we have experimentally verified that no fast dynamics could be observed in this configuration. This can be explained by $\hbar\omega$ being much larger than t .

Periodic driving. The periodic driving is implemented as in previous work²⁰. In brief, a piezo-electric actuator enables a controlled phase shift of the reflected X and \bar{X} lattice beams with respect to the incoming beams. To access the driven regime, we modulate the lattice position by a sinusoidal movement of the mirror position for the retroreflecting lattice beam at frequency $\omega/(2\pi)$. We choose the modulation to be along the direction of the horizontal bonds such that $V(x, y, z, \tau) \equiv V(x - A\cos(\omega\tau), y, z)$. We linearly ramp up a sinusoidal modulation and then maintain a fixed displacement amplitude A . During the modulation we ensure the correct phase relation $\varphi = 0.0(1)\pi$ between the two interfering X and Z lattice beams by modulating the phase of the respective incoming beams at the same frequency using acousto-optical modulators. In addition, this phase modulation is used to calibrate the phase and amplitude of the mirror displacement. In our set-up, the piezo modulation also leads to a residual periodic reduction in the interference amplitude of the lattice by at most 2%. For the lattice configurations used in our experiments, this shifts the mean tunnelling energy t_x down by about 2.5% and introduces a modulation of the tunnelling energy at twice the driving frequency $2\omega/(2\pi)$ with an amplitude of $\delta t = 0.025t_x$. The effect of the modulation is negligible because its amplitude has to be compared to the driving frequency. The effective driving strength is $\delta t/(\hbar\omega)$, which is always less than 3×10^{-3} in our case. In addition, we have verified that our experimental findings are not affected by the launching phase of the drive. The amplitude of the lattice displacement A is related to the normalized driving amplitude directly: $K_0 = m\omega d_x/\hbar$, where d_x is the distance between the two sites along the x direction. For our lattice potential, $d_x \neq \lambda/2$ and must be calculated for each individual configuration. To this end, we determine the Wannier functions located on the left and right sides of the bond considered, which are derived as the eigenstates of the band-projected position operator. The distance d_x is then evaluated as the difference between the eigenvalues of two neighbouring Wannier states, and is given in Extended Data Table 1 for all lattice configurations. In addition, because the lattice geometry in the x – z plane is not an ideal brick configuration, the bonds connecting two sites in the z direction are also slightly affected by the drive. The effective driving strength can be determined by the projected bond length on the modulation direction, which for our case is the x displacement $d_x^{\text{vert}} = \lambda/2 - d_x$ between neighbouring sites in the vertical z direction. The modulation amplitude is then $K_0^{\text{vert}} = d_x^{\text{vert}}/(d_x K_0)$. The values for d_x^{vert} are given in Extended Data Table 1 for our lattice configurations.

Calibration of the on-site interactions. The extension of the Wannier function can be similar to the scattering length for strong interactions in the optical potentials realized in our measurements. Thus, the actual on-site interaction strength U may be altered compared to the value calculated by using the non-interacting Wannier functions, as observed in previous experiments^{20,33}. We therefore determine U experimentally by driving the lattice at a frequency $\omega/(2\pi)$ and measure the number of double occupancies as a function of U . Double occupancies are maximally created either for $\hbar\omega = U$ in a connected lattice (Figs 2, 3) or for $\hbar\omega = (\sqrt{U^2 + 16t^2} + U)/2$ in the isolated double wells (Fig. 4). In the hexagonal lattice, the resonance position agrees within the uncertainty of the numerical value for U determined from the Wannier function, as shown in Fig. 3a. However, a substantial difference is observed in the isolated double wells. To account for this effect, we parameterize U by $U(a) = \alpha a(1 - a/a_c)$, where α is given by the non-interacting Wannier functions and a_c is a higher-order correction that depends on the lattice depth. For the isolated double wells, we find $a_c = 4,800(300)a_0$, which leads to a reduction in U of about 10% with respect to the calculated value for the datasets shown in Fig. 4c. Accordingly, this correction is incorporated into all interaction strengths given for the isolated double wells.

Validity of tight-binding approximation and higher band effects. When deriving the tight-binding Hamiltonian of the driven Fermi–Hubbard model in equation (1),

we assume that the Wannier functions are not modified by the modulation. However, for large driving amplitudes a substantial tilt is applied to the lattice in the co-moving frame, which introduces an energy bias $\hbar\omega K_0$ between neighbouring sites (see also Fig. 1). As a result, the Wannier functions will be modified by the admixture of higher-band Wannier functions of the untilted lattice. This will in turn lead to different tight-binding parameters t_x and U at any given time within the modulation cycle. To estimate the corrections that result from the change in the Wannier functions, we consider a cut through the tilted lattice potential in the x direction of the modulation. This potential can be very well approximated around the horizontal bonds by a lattice with a relative phase $\theta \neq \pi$ between the lattice beams \bar{X} and X (see equation (4)). The approximation in this step is to assume that all lattice sites in a given sublattice (A or B) are at equal energy. This is well justified for our lattice geometry because the tunnelling energy across the hexagon is zero and so the Wannier functions on the A sublattice, for example, are not influenced by the B sites to their left. Because the discrete spatial periodicity is restored in the lattice potential with $\theta \neq \pi$, we can compute the Wannier functions for any given energy bias and calculate the corresponding tight-binding parameters. The modulated lattice potential can then be described by a tight-binding Hamiltonian as in equation (1), where in addition to the oscillating force $f(\tau)$ the Hubbard parameters $t_x(\tau)$ and $U_{A,B}(\tau)$ become time- and sublattice-dependent. We decompose the parameters into their Fourier components, which take the form

$$\begin{aligned} t_x(\tau) &= t_x(K_0=0) + \delta t_0(K_0, \omega) + \delta t_2(K_0, \omega) \cos(2\omega\tau) + \dots \\ U_A(\tau) &= U(K_0=0) + \delta U_0(K_0, \omega) + \delta U_1(K_0, \omega) \cos(\omega\tau) \\ &\quad + \delta U_2(K_0, \omega) \cos(2\omega\tau) + \dots \\ U_B(\tau) &= U_A(\tau + \pi/\omega) \end{aligned}$$

The expansion of $t_x(\tau)$ features only even harmonics of ω because $t_x(\tau) = t_x(\tau + \pi/\omega)$. The main effect of the modulation is a shift in the static tunnelling energy by $\delta t_0(K_0, \omega)$, which is given in Extended Data Table 1 for the maximum driving amplitude and frequency in each lattice configuration. Note that even though the relative change in the tunnelling energy is around 10%–20% for large values of K_0 , the absolute change is much smaller because the hopping amplitude is renormalized by the Bessel function $\mathcal{J}_0(K_0)$ or $\mathcal{J}_1(K_0)$, depending on the frequency regime. On the other hand, we find that the shift in the mean value of U is much smaller, and even for the strongest driving we have $\delta U_0(K_0, \omega)/U < 6 \times 10^{-3}$. The second effect is a modulation of t_x and U , which is negligible because it has to be compared to the driving frequency. The dimensionless modulation strength for the lowest Fourier components will be given by $K_0^t = \delta t_2(K_0, \omega)/(2\hbar\omega)$ and $K_0^U = \delta U_1(K_0, \omega)/(\hbar\omega)$. Even for the maximum values of K_0 and ω , we find $K_0^t < 6 \times 10^{-3}$ and $K_0^U < 0.02$ for all of our lattice geometries. We also performed a numerical simulation of the two-site Hubbard model including all of the above modifications, in which we use a Trotter decomposition to evaluate the quasi-energy spectrum (see also Methods section ‘Theoretical treatment of the driven double well’ and Extended Data Fig. 4). We have found that even for the largest driving amplitudes used in the measurement of the exchange energy (see Fig. 4), J_{ex} is modified by at most 10 Hz in the off-resonant driving regime (compare to Extended Data Fig. 4b) and 60 Hz in the near-resonant case (Extended Data Fig. 4d), which is caused mainly by the shift in the mean value of t_x . This change is still smaller than or comparable to the uncertainty on the exchange energy that results from an imprecise calibration of the Hubbard parameters in the lattice, which is around 70 Hz.

Measurement of magnetic exchange. The exchange energy is measured in a Ramsey-type protocol in isolated double wells. After preparing singlet states on adjacent sites in a deep cubic lattice with $V_{\bar{X},X,\bar{Y},Z} = [30, 0, 30, 30]E_R$ as outlined above, we perform a $\pi/2$ pulse with a magnetic-field gradient to generate a coherent superposition between the singlet and triplet states. After this, we first ramp the magnetic field, the interfering lattice V_X and the driving amplitude K_0 to the desired value within 2 ms. In the next step, we trigger an exchange oscillation by suddenly lowering the barrier in the double well by decreasing $V_{\bar{X}}$ to the desired value within 100 μs . After a variable evolution time τ_{evol} in the driven system, we freeze the dynamics again by increasing $V_{\bar{X}}$ to $30E_R$ within 100 μs , revert the ramps of the magnetic field, the interfering lattice V_X and the driving amplitude K_0 , and perform a second $\pi/2$ pulse with a magnetic-field gradient. Finally, we measure the fraction of singlet states on adjacent sites, which after the evolution is $p_s(\tau_{\text{evol}}) = [1 - \cos(J_{\text{ex}}\tau_{\text{evol}}/\hbar)]/2$. In the experiment, we vary the evolution time τ_{evol} and measure the singlet fraction for each modulation amplitude K_0 for not fewer than 9 different values of τ_{evol} , with at least 27 individual measurements in total. We fit the data with a function $p_s(\tau_{\text{evol}}) = \alpha[1 - \cos(J_{\text{ex}}\tau_{\text{evol}}/\hbar)]\exp(-\beta\tau) + \gamma$ and extract the exchange from the fitted frequency. To estimate the error, we use a

resampling method that assumes a normal distribution of measurement results at each evolution time. The standard deviation of the distribution is determined by the measured standard deviation or, if we measured the singlet fraction at this τ_{evol} only once, by the residual from the fitted curve. Afterwards, we randomly sample a value for the singlet fraction at each evolution time and refit the resulting dataset. At the same time, the initialization values of the fit parameters J_{ex} and β are varied by $\pm 10\%$. This procedure is repeated 1,000 times and the mean \pm standard deviation of the resulting distribution of frequencies determines the asymmetric error bars for the fitted exchange frequency, as shown in Fig. 4. To demonstrate the sign change of the magnetic exchange for $U \lesssim \hbar\omega$ (Fig. 4d), we first let the system evolve for a time τ_0 with a non-driven exchange $J_{\text{ex}}^{(0)}$ until a quarter exchange oscillation has been performed, that is, $J_{\text{ex}}^{(0)}\tau_0 = \pi/2$. After that, we suddenly switch on the sinusoidal modulation at the desired value of K_0 , which projects the system onto a Hamiltonian with a negative J_{ex} . Therefore, the system changes its sense of rotation on the Bloch sphere (Fig. 4a) and the singlet fraction after a variable total evolution time $\tau_{\text{evol}} > \tau_0$ is given by $p_s(\tau_{\text{evol}}) = \{1 + \text{sgn}(J_{\text{ex}}) \times \sin[|J_{\text{ex}}|(\tau_{\text{evol}} - \tau_0)/\hbar]\}/2$.

Theoretical treatment of the driven double well. We perform both analytic and numerical studies on the driven double well, as described in earlier work²⁰. In this context, we use Floquet’s theorem to derive an effective static Hamiltonian in a high-frequency expansion. In the following, we include terms up to order $1/\omega$, as given in appendix A in ref. 20. In the off-resonant case, the term proportional to $1/\omega$ vanishes, such that the effect of the modulation is a pure renormalization of the tunnelling by a zeroth-order Bessel function $t \rightarrow t\mathcal{J}_0(K_0)$. Therefore, the exchange energy defined as the energy difference between the triplet and singlet state becomes

$$J_{\text{ex, off-res}} = \frac{1}{2} \left[-U + \sqrt{16t^2\mathcal{J}_0^2(K_0) + U^2} \right]$$

In the Heisenberg limit of large interactions ($t \ll U \ll \hbar\omega$) we find

$$J_{\text{ex, off-res}} \xrightarrow{U \gg t} 4 \frac{t^2}{U} \mathcal{J}_0^2(K_0) \quad (5)$$

In the case of near-resonant driving ($t \ll U \approx \hbar\omega$), we can express the Hamiltonian in terms of t , U and the detuning $\delta = \hbar\omega - U$, and we consider terms up to orders $\mathcal{O}(t^2/U, t\delta/U, \delta^2/U)$. In this regime, the single-particle tunnelling $t_0 = t\mathcal{J}_0(K_0)$ is renormalized as for the off-resonant case. On the other hand, the density-assisted tunnelling that changes the number of double occupancies is given by $t_1 = t\mathcal{J}_1(K_0)$. The exchange for $\delta \geq 0$ is given by

$$J_{\text{ex, res}} = \frac{1}{2} \left[\delta + 4 \frac{t_0^2}{U} \mp \sqrt{16t_1^2 + \left(\delta - 4 \frac{t_0^2 + t_1^2}{U} \right)^2} \right]$$

which reproduces the Heisenberg limit in equation (5) for the case of no driving ($K_0 = 0$). For large detunings ($t \ll \delta \ll U, \hbar\omega$), the exchange takes the form

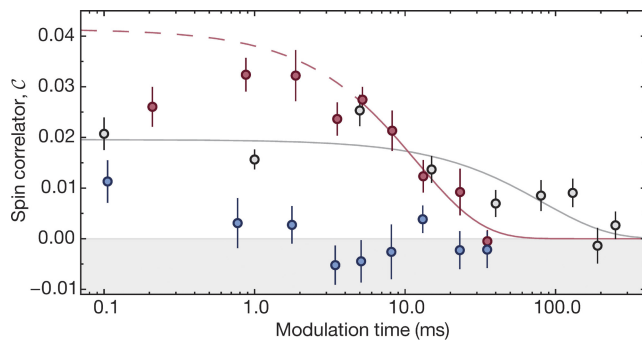
$$J_{\text{ex, res}} \xrightarrow{\delta \gg t} -4 \frac{t_1^2}{\delta} + 2 \frac{2t_0^2 + t_1^2}{U}$$

The leading term of this expansion is proportional to $\mathcal{J}_1^2(K_0)$ and changes sign with the detuning δ . This explains the switch to a ferromagnetic exchange for $U < \hbar\omega$ beyond a certain driving strength. In addition to the analytic derivation of the effective Hamiltonian, we also perform a numerical simulation of the two-site Hubbard model. We use a Trotter decomposition to evaluate the evolution operator over one period, from which we extract the spectrum (for details see ref. 20). A comparison of the numerical and analytic results for the experimental parameters is shown in Extended Data Fig. 4. For all of the derivations above, we assume that the static double well can simply be described by the tunnelling t and the on-site interaction U . However, if the Wannier functions on the two sites have a substantial overlap, then the description needs to be extended to a two-band Hubbard model. In this case, higher-order corrections such as density-assisted tunnelling δt , as well as nearest-neighbour interactions, direct exchange and correlated pair tunnelling V (the last three are all equal for the two-band Fermi–Hubbard model), become important (see appendix A.1 in ref. 20). For the experimental parameters in the off-resonant case (Fig. 4b), the values of these higher-order corrections are $V/\hbar = 2.4(7)$ Hz and $\delta t/\hbar = 22(3)$ Hz in the static lattice. In the near-resonant driving regime (Fig. 4c), interactions are stronger and the corrections increase to $V/\hbar = 26(8)$ Hz and $\delta t/\hbar = 120(10)$ Hz for $U/\hbar = 6.5(1)$ kHz, and

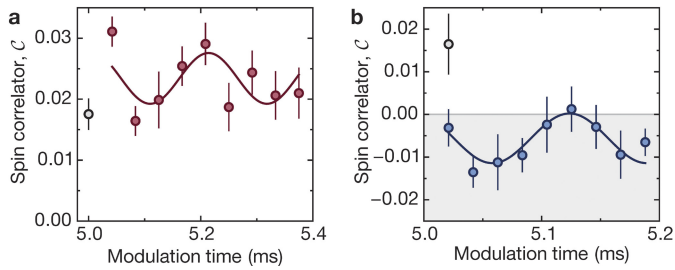
$V/h = 40(10)$ Hz and $\delta t/h = 170(20)$ Hz for $U/h = 9.1(1)$ kHz. To lowest order, the density-assisted tunnelling will increase the effective tunnelling to $t + \delta t$, and V decreases the exchange interaction by $2V$, in both the static and driven cases.

Data availability. All data files are available from the corresponding author on request. Source Data for Figs 2–4 and Extended Data Figs 1–3 are provided with the online version of the paper.

31. Tarruell, L., Greif, D., Uehlinger, T., Jotzu, G. & Esslinger, T. Creating, moving and merging Dirac points with a Fermi gas in a tunable honeycomb lattice. *Nature* **483**, 302–305 (2012).
32. Greif, D., Uehlinger, T., Jotzu, G., Tarruell, L. & Esslinger, T. Short-range quantum magnetism of ultracold fermions in an optical lattice. *Science* **340**, 1307–1310 (2013).
33. Uehlinger, T. *et al.* Artificial graphene with tunable interactions. *Phys. Rev. Lett.* **111**, 185307 (2013).

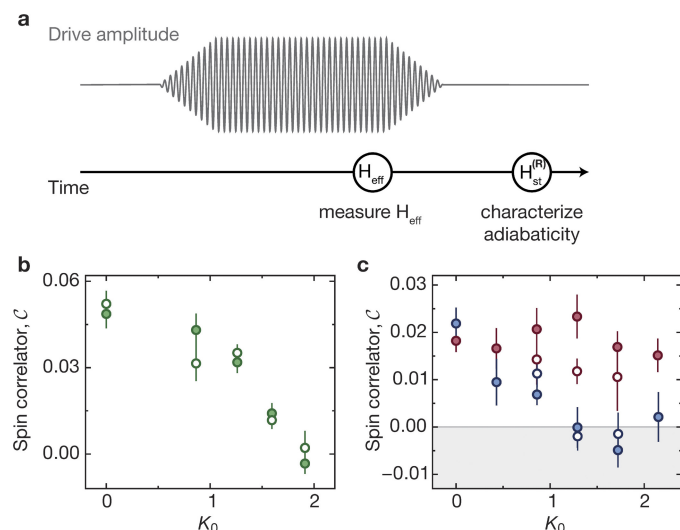


Extended Data Figure 1 | Time dependence of magnetic correlations for near-resonant driving. Nearest-neighbour spin–spin correlator C for the same lattice configuration as in Fig. 3, as a function of the modulation time after the ramp up of the drive. The data allow us to compare the formation and decay of magnetic correlations for two specific sets of interactions and modulation frequencies with the level of correlations in the static case (black). For a driving strength of $K_0 = 1.30(3)$ and with $U/\hbar = 3.8(1)$ kHz and $\omega/(2\pi) = 3$ kHz (red), antiferromagnetic correlations increase with time and reach a level higher than the static case (black, $U/\hbar = 3.8(1)$ kHz). If the interaction is smaller than the driving frequency (blue, $U/\hbar = 4.4(1)$ kHz, $\omega/(2\pi) = 6$ kHz), then the correlations switch sign and become ferromagnetic after a few milliseconds. For long times, the correlations in each configuration decrease as a result of heating in the lattice. Solid lines show exponential fits of the full data in the static case (grey) and to modulation times longer than 4 ms in the driven lattice for $U > \hbar\omega$ (red). The difference between the data and the dashed component of the fit (red) indicates an initial increase in the correlations. The extracted lifetimes decrease from 82(34) ms without drive to 12(4) ms at $K_0 = 1.30(3)$. All measurements are averaged over one modulation cycle. Data points and error bars denote the mean and standard error of 13 individual measurements at different times within one driving period (see Methods).



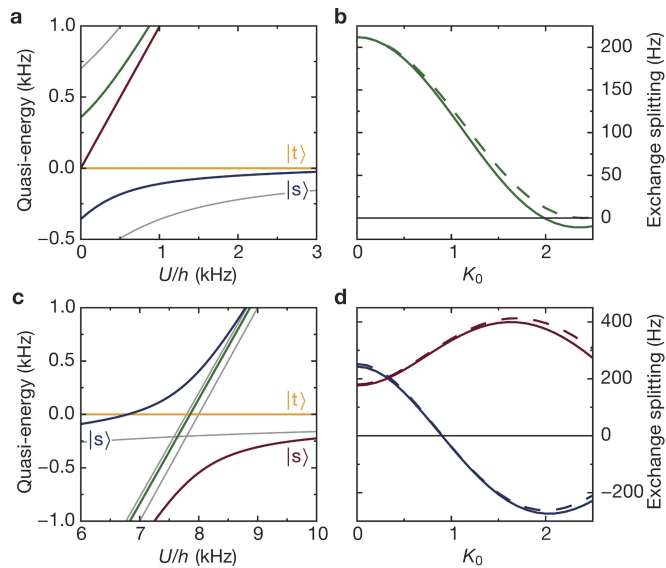
Extended Data Figure 2 | Micromotion for near-resonant driving.

a, b, Nearest-neighbour spin–spin correlator C for the lattice configuration in Fig. 3 and $K_0 = 1.30(3)$, as a function of modulation time after the ramp up of the drive, sampled within one oscillation period. We observe substantial micromotion both for the case of enhanced antiferromagnetic correlations (**a**; $U/h = 3.8(1)$ kHz and $\omega/(2\pi) = 3$ kHz) and for ferromagnetic correlations (**b**; $U/h = 4.4(1)$ kHz and $\omega/(2\pi) = 6$ kHz). For a different set of parameters in the measurement of the micromotion it should be also possible to switch between antiferromagnetic and ferromagnetic correlations within one driving cycle. The open symbols represent a reference measurement in the static case with all other parameters being equal. Solid lines are sinusoidal fits to the data, which results in a fitted frequency of $4.8^{+1.9}_{-0.4}$ kHz (**a**) or $7.6^{+3.9}_{-1.7}$ kHz (**b**). Error bars denote the standard error of 10 independent measurements.



Extended Data Figure 3 | Adiabaticity of the modulation ramp in the many-body system.

a, Starting from the static lattice, the modulation amplitude is ramped up and subsequently kept at a fixed value to allow for a 5 ms equilibration time. The ramp up time depends on the chosen configuration and is 3.333 ms (2 ms) for a modulation frequency of $\omega/(2\pi) = 3$ kHz (6 kHz). We start the detection of nearest-neighbour spin-spin correlations \mathcal{C} in the effective Hamiltonian H_{eff} by quenching the tunnelling to zero as we ramp up the lattice depth in all directions during the modulation within 100 μs . To estimate the adiabaticity of the final state, we perform a second type of measurement in which we revert the driving ramp and subsequently wait an additional 5 ms before the detection in the reverted static Hamiltonian $H_{\text{st}}^{(R)}$. If the ramp scheme of the modulation is fully adiabatic, we expect a reversal of the correlations to their static value. **b**, The nearest-neighbour spin-spin correlator \mathcal{C} is plotted against the modulation amplitude in the off-resonant driving regime ($U/\hbar = 0.93(2)$ kHz, $\omega/(2\pi) = 6$ kHz). The filled green circles are measured in the modulated system (same data as in Fig. 2b) and the open green circles after ramping off the modulation. The correlations no longer reach the level of the static case at $K_0 = 0$ after reverting the ramp. We attribute this to some extent to a reduced lifetime of correlations, which is found to be 14(5) ms at $K_0 = 1.26(4)$, compared to 92(16) ms in the static case. **c**, Spin-spin correlator for different driving strengths K_0 in the near-resonant regime for $U < \hbar\omega$ (blue; $U = 4.4(1)$ kHz, $\omega/(2\pi) = 6$ kHz) and in the regime of enhanced antiferromagnetic correlations (red; $U/\hbar = 3.8(1)$ kHz, $\omega/(2\pi) = 3$ kHz). Filled data points represent the effective states in the modulated system and open data points are measured after ramping off the modulation. Again, correlations do not reach the static value after reverting the driving ramp, owing to the finite lifetime (see also Extended Data Fig. 1). Data points and error bars denote the mean and standard error of 10 individual measurements at different times within one driving period (see Methods).



Extended Data Figure 4 | Analytical and numerical treatment of a driven double well. **a**, Quasi-energy spectrum for two particles in a double well as a function of the onsite interaction U for off-resonant driving ($t/h = 350$ Hz, $K_0 = 1.5$, $\omega/(2\pi) = 8$ kHz). Each of the four Floquet states representing the quasi-energy spectrum is shown in a distinct colour. The grey lines show the energy spectrum without modulation. For $U \gg t$, the ground state is the spin singlet $|s\rangle$ and the first excited state is the triplet $|t\rangle$. To lowest order, the driving renormalizes the tunnelling by a zeroth-order Bessel function $t_x \rightarrow t_x^{\text{eff}}(K_0) = t_x \mathcal{J}_0(K_0) \approx 0.51 t_x$. **b**, Calculated exchange energy $J_{\text{ex,off-res}}$ (see Methods), defined as the energy difference between the spin singlet and triplet states (see **a**), as a function of the driving amplitude K_0 for an off-resonant modulation ($t/h = 350$ Hz, $U/h = 2.1$ kHz, $\omega/(2\pi) = 8$ kHz; compare with Fig. 4b). The dashed line is the analytical result derived from a high-frequency expansion of the effective Hamiltonian; the solid line is the result of a numerical calculation. The exchange energy is reduced to small values as the tunnelling is renormalized by the zeroth-order Bessel function $\mathcal{J}_0(K_0)$. For large modulation amplitudes, deviations from the result obtained from an expansion up to order $1/\omega$ can be observed. Here, the exchange already becomes weakly ferromagnetic owing to the finite value of the interaction. **c**, Floquet spectrum of the double-well system as a function of the interactions U for near-resonant driving ($t/h = 640$ Hz, $K_0 = 0.8$, $\omega/(2\pi) = 8$ kHz). The grey lines show the energy spectrum without periodic modulation. The drive couples the singlet state to a state that contains double occupancy, which leads to an avoided crossing at $U \approx \hbar\omega$. As a result, a gap opens that is to lowest order given by $4\mathcal{J}_1(K_0)$. **d**, Dependence of the exchange energy $J_{\text{ex,res}}$ on the modulation amplitude in the near-resonant regime for two different detunings with $t/h = 640$ Hz and $\omega/(2\pi) = 8$ kHz (blue, $U/h = 6.5$ kHz; red, $U/h = 9.1$ kHz; compare with Fig. 4c). The dashed line is the analytical result (see Methods) derived from a high-frequency expansion of the effective Hamiltonian; the solid line is the result of a numerical calculation. For $U > \hbar\omega$ the exchange energy is greatly increased, whereas for $U < \hbar\omega$ it changes sign to ferromagnetic behaviour. In both driving regimes, the analytical result is in very good agreement with the numerics. Our measurements of the exchange energy in Fig. 4 agree well on a qualitative level with the theoretical expectation.

Extended Data Table 1 | Summary of experimental parameters

Main text figure	2	3	4b	4c, 4d
Atom number (10^3)	28(2)	32(2)	186(6)	
Initial T/T_F	0.07(1)	0.12(2)	0.06(1)	
$\bar{\omega}_{\text{trap}}/2\pi$ (Hz)	84(2)	84(2)	119(2)	
t_x/h (Hz)	810(150)	570(110)	350(50)	640(90)
t_y/h (Hz)	125(8)	125(8)	< 1	
t_z/h (Hz)	78(8)	85(8)	< 2	
$d_x/(\lambda/2)$	0.71(2)	0.74(2)	0.79(1)	0.73(1)
$d_x^{\text{vert}}/(\lambda/2)$	0.29(2)	0.26(2)	0.21(1)	0.27(1)
$\delta t_0(K_0^{\text{max}}, \omega^{\text{max}})/t_x$	0.085(1)	0.102(1)	0.236(9)	0.106(2)

Values given for Fig. 2 correspond to the initial static configuration with $K_0=0$. The initial temperature is measured before loading the atoms into the lattice. d_x is the length of the horizontal bonds; d_x^{vert} is the horizontal distance between two sites that form the vertical bonds in the z direction, which results from a non-rectangular unit cell. The effective modulation amplitude is given by the projection of each bond on the x direction. δt_0 describes the change in the mean value of t_i in the driven lattice due to a time-dependent modification of the Wannier functions. The values given here are upper bounds corresponding to the maximum modulation amplitude K_0^{max} and frequency ω^{max} used in each lattice configuration (see Methods for further details).

A photophoretic-trap volumetric display

D. E. Smalley¹, E. Nygaard¹, K. Squire¹, J. Van Wagoner¹, J. Rasmussen¹, S. Gneiting¹, K. Qaderi¹, J. Goodsell¹, W. Rogers¹, M. Lindsey¹, K. Costner¹, A. Monk¹, M. Pearson¹, B. Haymore¹ & J. Peatross²

Free-space volumetric displays, or displays that create luminous image points in space, are the technology that most closely resembles the three-dimensional displays of popular fiction¹. Such displays are capable of producing images in ‘thin air’ that are visible from almost any direction and are not subject to clipping. Clipping restricts the utility of all three-dimensional displays that modulate light at a two-dimensional surface with an edge boundary; these include holographic displays, nanophotonic arrays, plasmonic displays, lenticular or lenslet displays and all technologies in which the light scattering surface and the image point are physically separate. Here we present a free-space volumetric display based on photophoretic optical trapping² that produces full-colour graphics in free space with ten-micrometre image points using persistence of vision. This display works by first isolating a cellulose particle in a photophoretic trap created by spherical and astigmatic aberrations. The trap and particle are then scanned through a display volume while being illuminated with red, green and blue light. The result is a three-dimensional image in free space with a large colour gamut, fine detail and low apparent speckle. This platform, named the Optical Trap Display, is capable of producing image geometries that are currently unobtainable with holographic and light-field technologies, such as long-throw projections, tall sandtables and ‘wrap-around’ displays¹.

Optical Trap Display (OTD) image points can be seen from almost all angles because their radiation is not limited by a bounding aperture. By contrast, holographic image points are not visible unless they lie on a line that begins at a diffractive two-dimensional (2D) surface (or an image of that surface) and ends at the viewer’s eye. This limitation, described as ‘clipping’ or ‘vignetting’³, persists regardless of the composition, resolution or orientation of the hologram. The practical effect of clipping is that a hologram must be viewed like a television and not like a water fountain. That is, for a hologram of finite size, the best achievable in-plane viewing angle is 360° about the display surface. However, the maximum viewing angle around any individual image point is smaller than 360° and decreases rapidly as the image point moves away from the holographic display surface. By contrast, a free-space volumetric display provides an in-plane viewing angle of 360° around every image point at any depth. Clipping precludes almost all of the display geometries commonly associated with future three-dimensional (3D) displays, including long-throw projection, tall sandtables, and images that wrap around the viewer or other physical objects (Extended Data Fig. 1). These difficulties arise because holograms form points that are separate from the scattering surface. Conversely, volumetric displays may have scattering surfaces that are co-located with image points. The term ‘volumetric display’ is used to describe a device that “permits the generation, absorption, or scattering of visible radiation from a set of localized and specific regions within a physical volume” (ref. 4). The Display Technology Technical Group of the Optical Society of America has proposed⁵ a refinement of this definition, which specifies that a volumetric display has image points that are co-located with light scattering (or absorbing and generating) surfaces. This subtle distinction highlights how the sculpture-like physicality of volumetric

displays gives rise to their unique ability to present “depth rather than depth cues” (ref. 6). Among volumetric systems, we are aware of only three such displays that have been successfully demonstrated in free space: induced plasma displays^{7–9}, modified air displays^{10,11} and acoustic levitation displays¹². Plasma displays have yet to demonstrate RGB colour or occlusion in free space. Modified air displays and acoustic levitation displays rely on mechanisms that are too coarse or too inertial to compete directly with holography at present. The OTD advances the current state of the art by providing full-colour, free-space images with fine detail.

The OTD works by first trapping a micrometre-scale, opaque particle in a near-invisible (405-nm wavelength) photophoretic optical trap. The trapping sites are formed by a combination of oblique astigmatism and spherical aberration². Once a particle is confined, the trap is scanned, moving the particle through a volume in the space that it shares with the user. A system of collinear RGB lasers then illuminates the trapped particle to create a highly saturated, full-colour, low-speckle 3D image in space by persistence of vision (POV), as shown in Fig. 1a, c. The resulting images can have image points smaller than ten micrometres and can be seen from every angle (with the possible exception of the line forming the optical axis).

The display reported here is based on photophoretic optical particle trapping. Photophoretic traps are especially useful for confining and manipulating micrometre-diameter absorbing particles^{13–15}. Instead of radiation pressure and gradient forces, which are used by optical tweezer traps, photophoretic traps are thought to use thermal forces from ‘thermal accommodation’ and the radiometer effect. Forces arise on a particle because of uneven heating and thermal creep. Higher momentum is imparted to the particle from its hot side, leading to a net force pointing away from the heated region. A mathematical treatment of photophoretic effects for particles of different sizes is given in refs 16, 17. Our early trials were conducted using photophoretic traps that hold absorbing particles with greatly varying shapes and average diameters ranging from below 5 µm to above 100 µm—much larger than the mean free path of gas molecules (68 nm) at standard pressure and room temperature¹⁸. In this ‘continuous’ regime, the photophoretic force on a spherical particle is given as:

$$F_{\text{cont}} = \frac{3\pi}{2} \frac{\eta^2 d R \nabla T}{pM}$$

where R is the gas constant, η the viscosity of the gas, M the molecular weight of the gas, p the gas pressure, ∇T the gradient of the temperature T and d the diameter of the particle (from ref. 16). Several trap morphologies are possible, including optical bottle beams², optical vortices¹⁹, high-order doughnut beams²⁰ and Poisson spots. The traps used in this work are aberration traps, similar to the spherical aberration trap reported in ref. 2, and combine both spherical aberration and oblique astigmatism. By tilting the sagittal lens to add variable astigmatism to the fixed spherical aberration, tunable regions of high and low optical intensity were created near the lens focus. Two such traps were created for this work, one operating in the geometric optics regime

¹Department of Electrical and Computer Engineering, Brigham Young University, Provo, Utah 84602, USA. ²Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, USA.

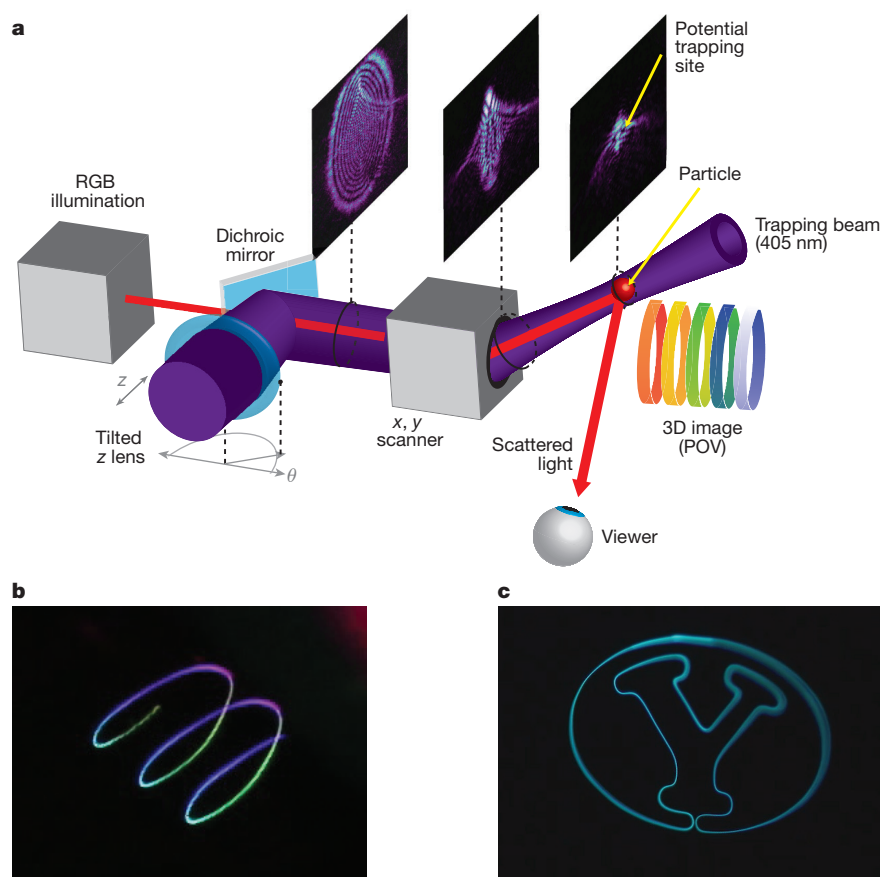


Figure 1 | The Optical Trap Display. **a**, A low-visibility light traps a particle and uses it to scan a volume. The resulting levitated optomechanical system is illuminated by an RGB laser source that can vary the illumination power and position of the trapped particle. As the particle scans the volume, images are formed by the POV method.

b, Photograph of an early OTD image. **c**, POV image. The particle in this image was scanned quickly enough (10 Hz) to produce an image by POV. A video of this OTD image is provided in Supplementary Video 1. Y logo from Brigham Young University Communications.

and the other in the diffraction mode. In the geometric trap, a low-visibility 405-nm-wavelength beam was passed through a lens tilted at 1° from the optical axis. Cross-sections and trapping sites for this beam are shown in Fig. 1a. This geometric instantiation system has the advantages that it is light, efficient and straightforward to implement. This trap was also produced with a liquid-crystal-on-silicon (LCOS) spatial light modulator (SLM), as described in Extended Data Fig. 2.

The most critical display parameters are summarized in Table 1. These parameters were collected from multiple prototypes (see Methods for additional details). Simple images have been demonstrated

at POV rates above the ten frames per second that are necessary for POV²¹. Figure 1c shows a vector image (1,307 vertices) traced at 12.8 frames per second (see Supplementary Video 1), which corresponds to 16,700 points per second, close to the maximum scanning rate of galvanometers (20,000 points per second). The volumetric image showed no noticeable flicker when viewed with the naked eye. An image of this complexity requires a particle velocity of 164 mm s^{-1} . Tests optimized for velocity (free from the processing delays and latency involved in image formation) show maximum achievable linear velocities greater than $1,827 \text{ mm s}^{-1}$, which suggests that it should be possible to obtain an order of magnitude increase in scan rate or image complexity for POV images without further optimization of either the trap or particle parameters. Supplementary Video 2 shows various particle movements including high accelerations. Accelerations in excess of $5g$, where g is the acceleration due to gravity, have been measured.

Our early observations indicate that display performance depends strongly on the quality of the optical trap parameters. Several automatic tests were run to identify the sensitive parameters for particle trapping. The trapping beam power, wavelength and numerical aperture are the parameters that have shown the greatest influence on trap quality, as manifested by hold time and airflow tolerance. High-contrast traps appear to hold best. Higher beam power is correlated with better trapping until the particle begins to disintegrate. Shorter wavelengths are associated with better trapping for both black liquor (cellulose) and tungsten particles. To give an example of common test parameters, a test consisting of 67 attempts and using 532-nm light at a power of 3.0 W has a pickup rate of 87% with an average hold time of 1.1 h. The maximum hold time recorded until now was in excess of

Table 1 | Prototype OTD parameters

Highest recorded linear speed until now	$1,827 \text{ mm s}^{-1}$
Highest image frame rate until now	12.8 frames per second, simple geometry (1,307 image points per frame), successful POV
Highest recorded acceleration until now	$57,574 \text{ mm s}^{-2}$ (5.67g)
Highest recorded hold time until now	17.2 h (measurement terminated by researcher)
Highest pickup rate until now	87% pickup success rate with average hold time of 1.1 h (sample size $N=67$)
Computational complexity	9 bytes per point per frame
Volume image resolution	Less than $10 \mu\text{m}$ minimum point dimension at all depths (around 1,600 dpi demonstrated)
Addressable volume	More than 100 cm^3
Colour characteristics	24-bit, laser-illuminated with no noticeable speckle
Scatter characteristics	Variable, in or out of plane, scatter angle variation from 360° to about 30°

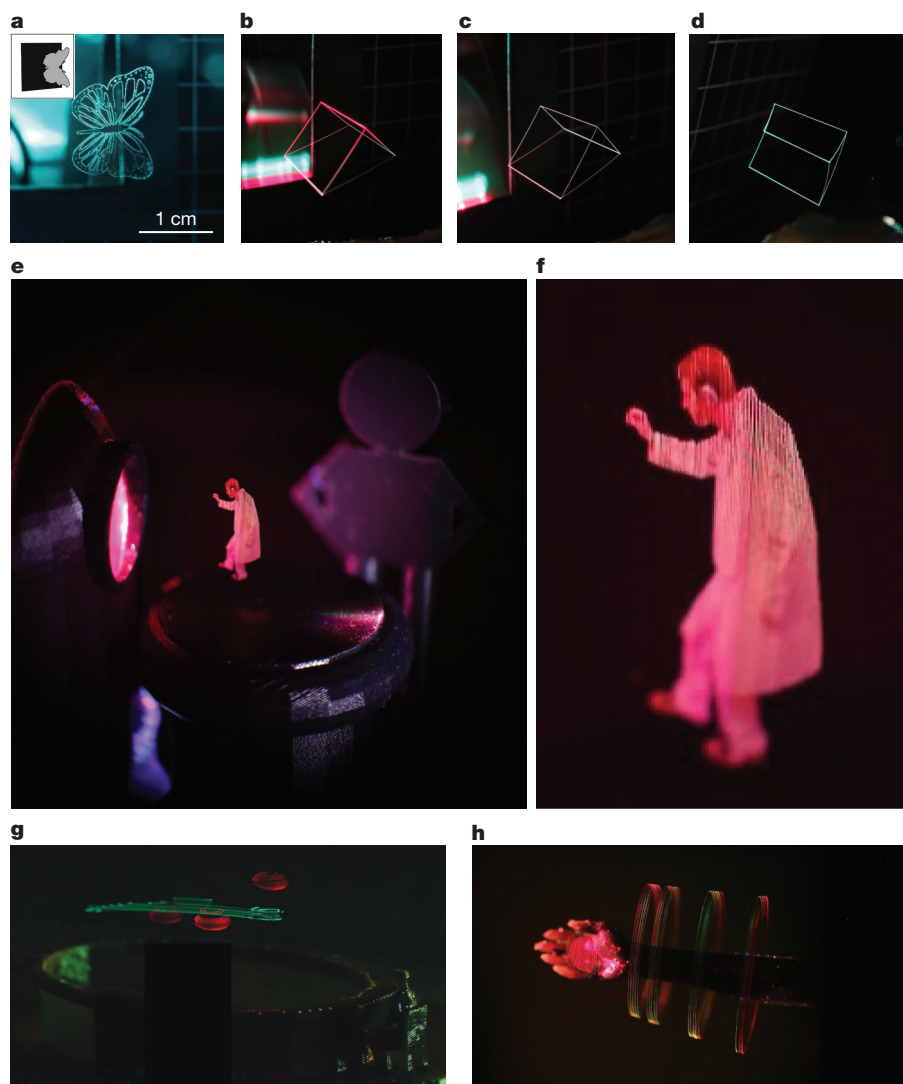


Figure 2 | 3D-printed light images produced by levitated optomechanics. **a**, Butterfly image, not clipped at the display aperture (butterfly image adapted from Rvector, Shutterstock). **b–d**, Prism viewable from all angles, including the side (**d**). **e**, ‘Projector’ geometry. The trap and illumination beams are emitted from the circular aperture (left) to form a projected image of one of the authors at a distance (centre).

f, Close-up of projected image. **g**, Tall sandtable. The 3D-printed sandtable does not clip the image above it. Model of sitting human figure from fletch55, 3D Warehouse. **h**, ‘Wrap-around’ image. Here vector rings surround a 3D-printed arm and a rastered circle illuminates the palm (arm model from clayguy, CGTrader). Image details, including exposure times, are given in Methods.

17.2 h, obtained at a wavelength of 532 nm and optical power of 3 W (the measurement was terminated by the researcher), and the minimum hold power recorded was less than 24 mW (for 405-nm light). Successful trapping has been performed with 635-nm, 532-nm, 445-nm and 405-nm light. Given that the particles have sizes of ten micrometres or higher, the resolution of the display is also determined by the addressable points of the scanning apparatus. The highest image resolution so far is 1,600 dots per inch (dpi) for a projected total of 5 billion addressable points in a pyramidal volume with a maximum linear dimension of two inches on each side and a depth of approximately one inch, which was achieved with the proposed instantiation technique.

When considering image quality, such as optical geometry, colour and resolution, no effort was made to optimize for write speed. Instead, the images in Figs 2 and 3 were created by long exposure (exposure times are listed in Methods).

The OTD was used to demonstrate its ability to bend the light path. Figure 2 shows examples of OTD images produced using levitated optomechanics to successfully create optical geometries unachievable by holograms. These OTD images are not clipped at the display

aperture (Fig. 2a), exhibit parallax and can be seen from all angles (Fig. 2b–d). The OTD can create long-throw projections (Fig. 2e, f), tall sandtables (Fig. 2g) and images that wrap around physical objects (Fig. 2h).

Colour was obtained from low-cost, commercial diodes emitting red, green and blue laser light propagating collinearly with the 405-nm trap beam. The diode sources were driven using 8-bit pulse-width modulation. Several colour tests were performed using the rastered images shown in Fig. 3. Figure 3a shows particles being illuminated by red, green and blue lasers. The colours are highly saturated, consistent with laser illumination. Figure 3b demonstrates the ability of the OTD to create additive colour and grayscale. Figure 3c shows an image frame with soft tones and no apparent speckle on an image of 3 cm by 2 cm. To demonstrate the resolution of the system, a 1-cm-diameter image of Earth was written above a fingertip at 1,600 dpi resolution (Fig. 3d).

The practical limitations imposed on the operation, scaling and complexity of OTD images include (i) finite mechanical scan speed, (ii) variable trapping conditions, (iii) airflow sensitivity and

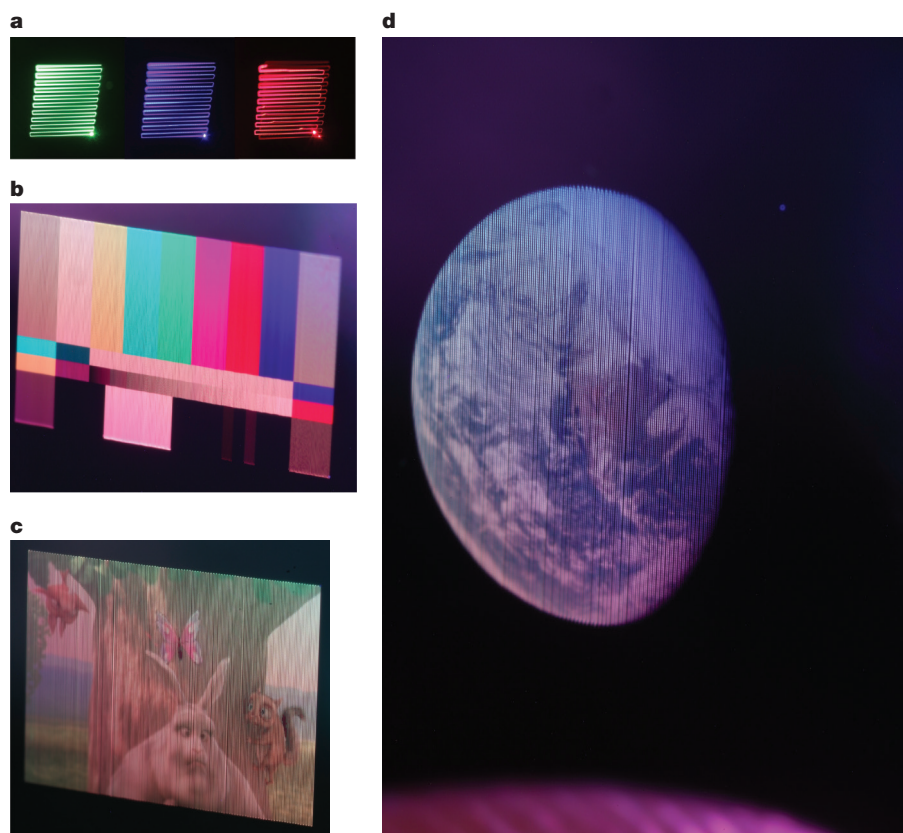


Figure 3 | Examples of the colour and resolution quality of the images. **a**, Laser-illuminated, trapped particles showing highly saturated RGB primaries. **b**, Additive colour combinations and grayscale. **c**, Full-colour image with soft tones and no discernible speckle at a scale of approximately $3\text{ cm} \times 2\text{ cm}$ (adapted from <http://www.bigbuckbunny.org>

under Creative Commons license CC BY 3.0, <https://creativecommons.org/licenses/by/3.0/>). **d**, Example of free-space Earth image above a human fingertip with a pixel dimension of approximately $10\text{ }\mu\text{m}$. The image was generated using NASA photograph number AC75-0027. Image details, including exposure times, are given in Methods.

(iv) illumination splash. Image complexity in a single-particle display is constrained primarily by the speed of the scanning system. This limitation may be overcome by the use of solid-state scanning and parallelism. Parallelism is achieved by simultaneously manipulating multiple trapped particles. That is to say, instead of having one particle responsible for all of the points in an image (one particle per image volume), multiple particles may be trapped, manipulated and illuminated independently using solid-state SLMs. The scanning requirements are reduced as more particles are added. A line of trapped particles reduces the drawing complexity to a two-axis scan (one particle per image plane), a 2D array of trapped particles reduces the scan complexity to a single-axis scan (one particle per image line), and a 3D array of trapped particles could eliminate the need for scanning entirely (one particle for every image point). SLM trapping and SLM particle manipulation have been used previously^{15,22–24}; we have successfully trapped particles using an LCOS SLM, with the phase pattern designed to create an astigmatic aberration trap (see Extended Data Fig. 2). As an example, an OTD with a single-axis scan (one particle per image line), operating at the current maximum linear velocity of $1,827\text{ mm s}^{-1}$, would be able to create images approximately 180 mm high at POV refresh rates (10 Hz).

The strength of particle trapping and holding varies greatly because of the wide distribution of the particle sizes and shapes, as well as the presence of multiple axial trapping sites of different sizes and quality. Under poor trapping conditions, a particle may hop from one trapping site to another. The maximum achievable particle velocity and acceleration seem to depend strongly on these highly variable trap conditions. A clearer upper bound on the complexity of single-particle images will be obtained when the optimal particle and trap morphologies are identified and isolated.

Particles are sensitive to airflow. Under good trapping conditions, trapped particles are robust to low levels of airflow, including airflow generated by human breathing and hand gestures (estimated²⁵ airflow upper bound of one litre per second). However, it is unlikely that the display would function outdoors without an enclosure unless particles were much more strongly confined or steps were taken to refresh trapped particles regularly.

In general, some of the illumination light does not scatter, but instead continues along the optical axis to form a laser ‘splash’. This can be overcome by carefully controlling the illumination focus, by directing the optical axis towards an absorbing surface or by using active particles. Examples of active particles are fluorescent particles or aerosol droplets trapped with an infrared beam and then illuminated with a low-power ultraviolet beam to induce coloured emission²⁶. p–n junctions might also be levitated and optically pumped to create high-gain, light-emitting particles with no visible splash.

This study provides a proof-of-concept demonstration of a full-colour, high-resolution, free-space volumetric POV display that can achieve display geometries beyond the capabilities of holography alone. The trapping strategy used in the OTD is both straightforward and efficient. The reported prototypes use commercial hardware and have low cost relative to other free-space volumetric displays. We anticipate that the device can readily be scaled using parallelism and consider this platform to be a viable method for creating 3D images that share the same space as the user, as physical objects would.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 August; accepted 6 November 2017.

1. McDonald, G. Why don't we have Princess Leia holograms yet? *All Tech Considered* <https://www.npr.org/sections/alltechconsidered/2017/08/26/546084473/why-don-t-we-have-princess-leia-holograms-yet> (2017).
2. Shvedov, V. G., Hnatovsky, C., Rode, A. V. & Krolikowski, W. Robust trapping and manipulation of airborne particles with a bottle beam. *Opt. Express* **19**, 17350–17356 (2011).
3. Blundell, B. G. On the uncertain future of the volumetric 3D display paradigm. *3D Res.* **8**, 11 (2017).
4. Blundell, B. G. & Schwarz, A. J. *Volumetric Three-Dimensional Display Systems* Vol. 1, 330 (Wiley-VCH, 2000).
5. Smalley, D. E. *OSA Display Technical Group Illumiconclave I* (Optical Society of America, 2015); <http://holography.byu.edu/Illumiconclave1.html>.
6. Clifton, T. & Wefer, F. L. Direct volume display devices. *IEEE Comput. Graph. Appl.* **13**, 57–65 (1993).
7. Kimura, H., Uchiyama, T. & Yoshikawa, H. Laser produced 3D display in the air. In *Proc. ACM SIGGRAPH 2006 Emerging Technologies* (Association for Computing Machinery, 2006).
8. Ochiai, Y. *et al.* Fairy lights in femtoseconds: aerial and volumetric graphics rendered by focused femtosecond laser combined with computational holographic fields. *ACM Trans. Graph.* **35**, 17 (2016).
9. Saito, H. *et al.* Laser-plasma scanning 3D display for putting digital contents in free space. *Proc. SPIE* **6803**, 680309 (2008).
10. Ruiz-Avila, J. Holovect: Holographic Vector Display. *Kickstarter* <https://www.kickstarter.com/projects/2029950924/holovect-holographic-vector-display> (2016).
11. Perlin, K. Volumetric display with dust as the participating medium. US patent 6,997,558 (2006).
12. Sahoo, D. R. *et al.* JOLED: a mid-air display based on electrostatic rotation of levitated Janus objects. In *Proc. 29th Annual Symposium on User Interface Software and Technology* 437–448 (ACM, 2016).
13. Davis, E. J. A history of single aerosol particle levitation. *Aerosol Sci. Technol.* **26**, 212–254 (1997).
14. Ehrenhaft, F. *Über die Messung von Elektrizitätsmengen, die kleiner zu sein scheinen als die Ladung des einwertigen Wasserstoffions oder Elektrons und von dessen Vielfachen abweichen* (Hölder, 1910).
15. Inaba, H., Sato, S. & Kikuchi, S. Method and apparatus for optical micro manipulation. US patent 5,363,190 (1994).
16. Horvath, H. Photophoresis—a forgotten force? *Kona Powder Part. J.* **31**, 181–199 (2014).
17. Rohatschek, H. Direction, magnitude and causes of photophoretic forces. *J. Aerosol Sci.* **16**, 29–42 (1985).
18. Jennings, S. The mean free path in air. *J. Aerosol Sci.* **19**, 159–166 (1988).
19. Desyatnikov, A. S., Shvedov, V. G., Rode, A. V., Krolikowski, W. & Kivshar, Y. S. Photophoretic manipulation of absorbing aerosol particles with vortex beams: theory versus experiment. *Opt. Express* **17**, 8201–8211 (2009).
20. He, H., Heckenberg, N. & Rubinsztein-Dunlop, H. Optical particle trapping with higher-order doughnut beams produced using high efficiency computer generated holograms. *J. Mod. Opt.* **42**, 217–223 (1995).
21. Bowen, R. W., Pola, J. & Matin, L. Visual persistence: effects of flash luminance, duration and energy. *Vision Res.* **14**, 295–303 (1974).
22. Galstian, T. Light-driven molecular rotational motor. US patent 6,180,940 (2001).
23. van Eymeren, J. & Wurm, G. The implications of particle rotation on the effect of photophoresis. *Mon. Not. R. Astron. Soc.* **420**, 183–186 (2012).
24. Wilson, S. D. & Clarke, W. L. Particle trap. US patent 5,170,890 (1992).
25. Gupta, J. K., Lin, C. H. & Chen, Q. Characterizing exhaled airflow from breathing and talking. *Indoor Air* **20**, 31–39 (2010).
26. Neukirch, L. P., Gieseler, J., Quidant, R., Novotny, L. & Vamivakas, A. N. Observation of nitrogen vacancy photoluminescence from an optically levitated nanodiamond. *Opt. Lett.* **38**, 2976–2979 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Baxter from Brigham Young University, Chemical Engineering Department, for use of his equipment, as well as G. Nielson, V. M. Bove Jr and P.-A. Blanche for discussions. We are grateful to S. Hilton for help with 3D printing. We acknowledge the Blender Foundation for content (Fig. 3c; adapted from <http://www.bigbuckbunny.org> under a Creative Commons license, <https://creativecommons.org/licenses/by/3.0/>) and software for 3D image conversion. We also acknowledge those who supplied source material for the holograms and volumetric images shown. B. Wilcox and Brigham Young University Communications provided footage of the Y logo in Fig. 1 and Supplementary Video 1. The butterfly image used to create the volumetric image and holograms shown in Fig. 2a and Extended Data Fig. 1a, b was sourced from Rvector, Shutterstock. The volumetric image in Fig. 3d was generated from NASA photograph number AC75-0027. The image of the arm shown in Fig. 2h and illustrated in Extended Data Fig. 1e was created using a model from clayguy, CGTrader. The model of the sitting human figure that is 3D-printed in Fig. 2g and illustrated in Extended Data Fig. 1d was sourced from fletch55, 3D Warehouse.

Author Contributions D.E.S. and J.P. formed the original concept. D.E.S. directed the research. E.N. developed the colour system and performed the analysis. K.S., K.C., B.H. and M.P. performed early trapping experiments and identified suitable particles. K.S., K.C., B.H., M.P. and J.V.W. developed the first monochrome system and the predecessor to the colour system. K.C. and J.R. developed the fast version of the monochrome system. J.R. developed the automatic pickup mechanism. S.G. and K.Q. developed the SLM trapping system. J.G. calculated the particle acceleration, performed the phase-mask analysis and conducted the literature review. K.C. and B.H. created the trapping chambers and performed aerosol experiments. W.R. created the hardware and software to migrate the system to full 3D images. M.L. created the raster-to-vector conversion algorithms. A.M. developed the driving architecture for full resolution images.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to D.E.S. (smalley@byu.edu).

Reviewer Information *Nature* thanks B. G. Blundell and C. Wang for their contribution to the peer review of this work.

METHODS

Colour display experiments were performed using an RGB laser system (OEM Laser Systems) combined with a 200-mW 405-nm laser using dichroic mirrors. The beam was expanded and focused through a tilted spherical lens mounted on an active translation stage, immediately followed by a 30-mm-aperture x/y galvanometric scanning assembly.

All images shown in this paper were made with long exposures, with the exception of Fig. 1c, which was written at a rate of greater than ten frames per second. For all other images, the speed of writing was constrained by the method of colour display, which paused at each pixel point to allow time for the green laser to reach full power. The resulting long-exposure image times were as follows: Fig. 2a, 16.2 s; Fig. 2b–d, 8.4 s; Fig. 2e, f, 40.8 s; Fig. 2g, 45.4 s; Fig. 2h, 56.4 s; Fig. 3b, 34.3 s; Fig. 3c, 18.9 s maximum; and Fig. 3d, 19 s. A portion of the 3D-printed surrounding scene in Fig. 2g was blacked out digitally.

The 532-nm frequency-doubled laser was swapped with a non-frequency-doubled, 520-nm diode laser with a 700-ns pulse generator for a faster modulation response time. The lasers were expanded to a beam waist of 30 mm and were caught by a positive, biconvex spherical lens tilted at an angle of 1° from the normal to the optical axis. The lens had a focal length of 125 mm. The x/y scanners had a maximum aperture of 30 mm with Al-coated mirrors. The galvanometric scanners were driven by a microcontroller. The focusing lens was mounted on a Physik Instrumente V-551.2B linear translation stage. The photographs in Figs 2 and 3 were taken through a 405-nm dichroic mirror that was used as an optical notch filter.

The early single-beam display prototypes were created using a 10-W Verdi laser operating at powers between 3 W and 4.06 W. The first instantiation system used a lens with 75-mm focal length and a single mirror constructed from an aluminized 100-mm silicon wafer, gimbal-mounted to scan over π steradians. This first instantiation system provided many of the results reported in this paper. An improved single-beam setup was created using a 125-mm lens and an OEM x/y scanning system with a 15-mm aperture. Early instantiation systems had both the trapping and illumination beams passing through the tilted lens. An iris was placed just before the lens to truncate the beam waist to improve pattern contrast at the focus. Particles were introduced into the trap either by scanning the trap into a retractable reservoir of particles or manually, by passing a plastic or metal instrument coated with particles away from the galvanometric scanner mirrors and through the focus. The retractable reservoir was composed of a piece of aluminium bar stock wrapped in aluminium foil. The aluminium foil was coated with a single layer of black liquor particles. The focused beam was pulled up and away from the aluminium surface to capture a particle. Once captured, the bar stock was pulled out of the write volume by a worm gear mechanism. The repeatability of trapping degraded gradually as a function of the number of pickup cycles. Efforts were made to reduce the airflow during trapping and image writing by placing foam core enclosures, baffles or table curtains around the write volume. Experiments were performed both within enclosures and in open air. Enclosures were found to improve trapping

and hold times in some circumstances. The most commonly trapped particle in the experiments reported here was black liquor with approximately 70% solid content. Black liquor is a cellulose solution that is a common by-product of the paper-making process. Samples used in this work were obtained from a kraft pulp mill using southern pine.

The LCOS experiments were performed using a Hamamatsu X10468-01 phase-only 792 pixel \times 600 pixel SLM. The device was illuminated by a 100-mW, 405-nm diode laser approximately 10° from the normal. Owing to beam clipping, which is required for uniform illumination and internal reflections, the power delivered to the trap was only 48 mW. The phase image displayed on the LCOS was created numerically by combining the spherical aberration, astigmatism and coma that can be created by lenses. This was then combined with the factory-provided correction phase image and an offset grating to create the final phase image.

By using a beam-shaping SLM, the dynamic functions of the OTD can be changed to those of a solid-state display (Extended Data Fig. 2). The trap was also produced with an LCOS SLM with a phase pattern described by the following relations:

$$P_{SA} = A_{SA} \rho^4$$

$$P_A = A_A \rho^2 (\cos^2 \theta - 1)$$

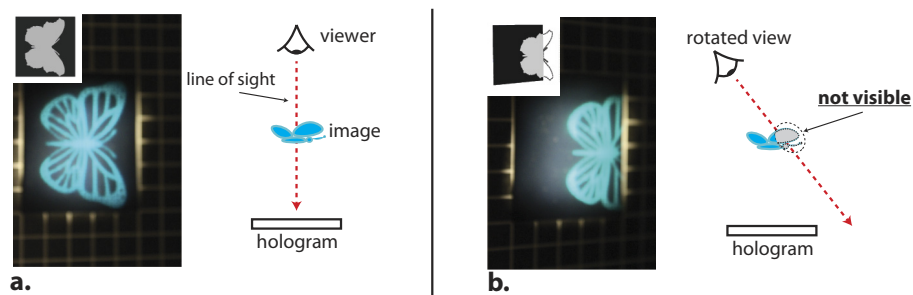
$$P_{IMG} = P_{SA} + P_A + P_{CAL}$$

where P_{SA} , P_A and P_{IMG} stand for the spherical aberration, astigmatic aberration and total phase images, respectively. P_{CAL} is a calibration phase image specific to the LCOS being used and is often provided by the manufacturer. The scalar coefficients A represent aberration weights (A_A , weight for astigmatic aberration; A_{SA} , weight for spherical aberration), θ describes the rotation of the lens from the perpendicular, and ρ is the radial distance from the centre of the LCOS display.

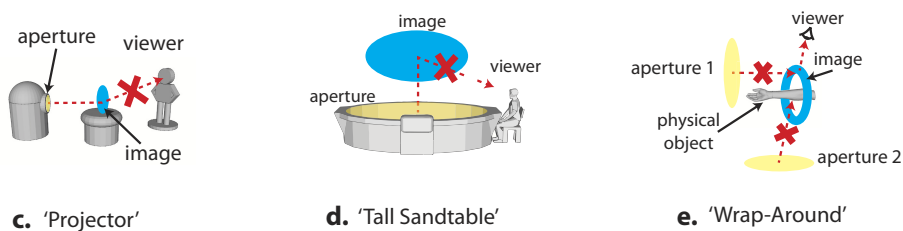
SLMs have previously been used for trapping and manipulating particles^{27,28}. To achieve parallel, independent trapping, a fixed phase plate or an active SLM may be used to trap multiple particles for scanning simultaneously. A combination of two SLMs or a phase plate and an SLM would provide independent trapping and illumination. We have confirmed that a particle can be held with a phase-only modulator displaying a diffraction pattern meant to both steer input light and modify the astigmatic and spherical aberration of the trap at the focus of the output lens.

Data availability. The data that support the findings of this study are available from the corresponding author on reasonable request.

27. Bowman, R. W. & Padgett, M. J. Optical trapping and binding. *Rep. Prog. Phys.* **76**, 026401 (2013).
28. Ashkin, A., Dziedzic, J. M. & Yamane, T. Optical trapping and manipulation of single cells using infrared laser beams. *Nature* **330**, 769–771 (1987)

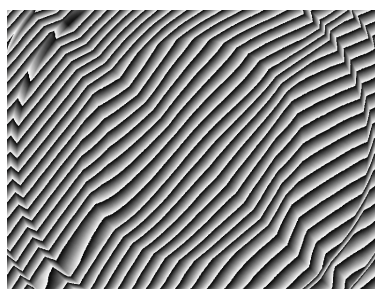


Forbidden Geometries

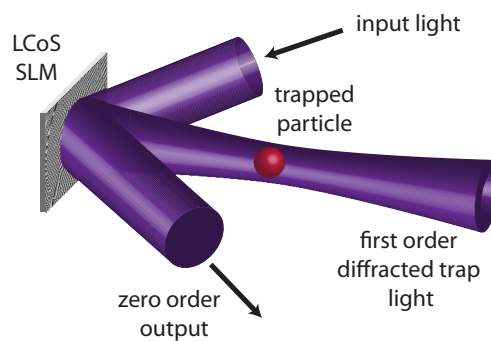


Extended Data Figure 1 | Limitations of holography. **a**, Holographic butterfly above hologram surface (butterfly image adapted from Rvector, Shutterstock). **b**, The holographic image is clipped at the display aperture when viewed from the side. **c**, Long-throw projector geometry. Light

projected from a holographic display aperture will not bend to reach the viewer in this geometry. **d**, **e**, The same is true for the tall sandtable geometry (**d**) and the wrap-around geometry (**e**). In **e**, a physical object (arm) obstructs the rays from aperture 2.



a.



b.

Extended Data Figure 2 | Solid-state astigmatic aberration trap. a. An LCOS pattern used to encode an aberration optical trap. **b.** The display trapping function can be changed to that of a solid-state display by encoding the phase pattern of the holographic trap on an SLM as shown.

Early episodes of high-pressure core formation preserved in plume mantle

Colin R. M. Jackson^{1,2}, Neil R. Bennett¹, Zhixue Du¹, Elizabeth Cottrell² & Yingwei Fei¹

The decay of short-lived iodine (I) and plutonium (Pu) results in xenon (Xe) isotopic anomalies in the mantle that record Earth's earliest stages of formation^{1–8}. Xe isotopic anomalies have been linked to degassing during accretion^{2–4}, but degassing alone cannot account for the co-occurrence of Xe and tungsten (W) isotopic heterogeneity in plume-derived basalts^{9,10} and their long-term preservation in the mantle. Here we describe measurements of I partitioning between liquid Fe alloys and liquid silicates at high pressure and temperature and propose that Xe isotopic anomalies found in modern plume rocks (that is, rocks with elevated ³He/⁴He ratios) result from I/Pu fractionations during early, high-pressure episodes of core formation. Our measurements demonstrate that I becomes progressively more siderophile as pressure increases, so that portions of mantle that experienced high-pressure core formation will have large I/Pu depletions not related to volatility. These portions of mantle could be the source of Xe and W anomalies observed in modern plume-derived basalts^{2–4,9,10}. Portions of mantle involved in early high-pressure core formation would also be rich in FeO^{11,12}, and hence denser than ambient mantle. This would aid the long-term preservation of these mantle portions, and potentially points to their modern manifestation within seismically slow, deep mantle reservoirs¹³ with high ³He/⁴He ratios.

Mantle plumes contain components with high ³He/⁴He ratios that have been broadly interpreted as evidence for a primordial reservoir within Earth^{1,14} (hereafter high-³He/⁴He mantle is referred to as 'plume mantle'). Recent high-precision analyses of Xe isotopes from plume mantle reveal a common signature of a low ¹²⁹Xe*/¹³⁶Xe*_{Pu} ratio compared to mid-ocean-ridge basalt (MORB) mantle^{2–4}. ¹²⁹Xe* is the decay product of short-lived and volatile ¹²⁹I. ¹³⁶Xe*_{Pu} is the decay product of short-lived and refractory ²⁴⁴Pu. Consequently, low ¹²⁹Xe*/¹³⁶Xe*_{Pu} ratios for plume mantle have been interpreted to reflect an early depletion of I related to degassing of materials accreted to Earth^{2,4}.

Other observations of plume mantle, however, are not readily explained by degassing processes. Modern plume mantle contains W isotopic anomalies, both higher and lower than MORB mantle^{9,10}, that result from the decay of short-lived hafnium (¹⁸²Hf). In contrast to I, both W and Hf are refractory and not affected by degassing. W anomalies reflect the separation of core-forming metal from the silicate mantle early in Solar System history owing to the strong tendency of the core to incorporate W and fractionate Hf/W ratios within terrestrial bodies^{15–18}. Moreover, early accretion of volatile-poor material, as would be required to generate low ¹²⁹Xe*/¹³⁶Xe*_{Pu} ratios related to I volatility, would result in FeO-poor, buoyant mantle¹¹, which would be difficult to sequester and preserve at the base of the mantle where plumes originate¹⁹. In this study, we test the hypothesis that Xe and W anomalies in plume mantle were generated by the same process—core formation at high pressures—by quantifying how I partitions between liquid Fe alloy and liquid silicate at equilibrium ($D_{\text{met/sil}}^{\text{I}} = [I]_{\text{met}}/[I]_{\text{sil}}$, where brackets denote concentration).

Limited experimental data at 2–20 GPa (about 2,800 K) demonstrate that I can be siderophile and that its partitioning behaviour is affected

by alloying components in liquid Fe²⁰, but the pressure, temperature and compositional (P – T – X) dependencies of $D_{\text{met/sil}}^{\text{I}}$ remain unknown under conditions directly relevant to deep magma oceans. If $D_{\text{met/sil}}^{\text{I}}$ values are sensitive to P – T – X under the conditions associated with core formation, then I/Pu variations should be formed with co-variations in Hf/W in the mantle. Over time, these co-variations of I/Pu and Hf/W would evolve coupled Xe and W anomalies if mixing with the remaining mantle was limited.

We define the P – T – X dependencies of $D_{\text{met/sil}}^{\text{I}}$ by conducting partitioning experiments in a laser-heated diamond anvil cell (DAC) (beam-line 13-ID-D at the Advanced Photon Source) at temperatures from 3,100 K to 4,900 K, pressures from 20 GPa to 45 GPa, and oxygen fugacity from -0.5 to -2.0 logarithmic units relative to the iron–wüstite buffer (Methods, Supplementary Table 1). These are conditions directly relevant to deep magma oceans and related core formation events (such as in ref. 11). Cross-sections of heating spots were exposed for chemical analysis using a focused ion beam (Fig. 1).

The concentration of I and other elements in metal and silicate phases were quantified using a JEOL 8530F microprobe (Methods). Measured $D_{\text{met/sil}}^{\text{I}}$ values (as atomic ratios) vary from 0.32 to 14 ($n = 18$, Table 1). To relate $D_{\text{met/sil}}^{\text{I}}$ variations in the present experiments to P – T – X conditions of core formation, we conducted a stepwise, unweighted least-squares regression (Methods, Extended Data Fig. 1, Supplementary Table 2):

$$\ln(D_{\text{met/sil}}^{\text{I}}) = (461 \pm 147)P/T - (12 \pm 3)X_{\text{met}}^{\text{S}} - (21 \pm 4)X_{\text{met}}^{\text{O}} - (5 \pm 1) - \ln(\gamma_{\text{met}}^{\text{Fe}}) \quad (1)$$

where

$$X_{\text{met}}^i = T_r/T \left[[i]_{\text{met}} \left(1 + \frac{\ln(1 - [i]_{\text{met}})}{[i]_{\text{met}}} - \frac{1}{1 - [i]_{\text{met}}} \right) - [i]_{\text{met}}^2 [I]_{\text{met}} \left(\frac{1}{1 - [I]_{\text{met}}} + \frac{1}{1 - [i]_{\text{met}}} + \frac{[I]_{\text{met}}}{2(1 - [I]_{\text{met}})^2} - 1 \right) \right] \quad (2)$$

for element $i = \text{O}$ or S , and $\gamma_{\text{met}}^{\text{Fe}}$ is the activity coefficient for Fe in the liquid alloy calculated by equation (23) of ref. 21, $[i]_{\text{met}}$ is the atomic fraction of S or O present in the metal alloy, and T_r is 1,873 K. Quoted uncertainties are 1σ standard errors. Negative coefficients for the X_{met}^i terms signify more siderophile partitioning behaviour because X_{met}^i correlates negatively with $[i]_{\text{met}}$. The magnitudes of the coefficients associated with the P/T and $X_{\text{met}}^{\text{S}}$ terms are corroborated by a series of piston cylinder experiments (Extended Data Fig. 2, Supplementary Tables 2 and 3).

Application of this parameterization to MORB mantle core formation models yields a $D_{\text{met/sil}}^{\text{I}}$ value of 3.4 ± 1.1 (1σ) (blue circle, Fig. 2a) and confirms that the core can be an important reservoir for bulk Earth I and ¹²⁹Xe* (ref. 20). This estimate for $D_{\text{met/sil}}^{\text{I}}$ is derived using the P – T – X conditions (38 GPa, 3,500 K) that produce mantle with bulk

¹Geophysical Laboratory, Carnegie Institution of Washington, Washington DC 20015, USA. ²Department of Mineral Sciences, National Museum of Natural History, Smithsonian Institution, Washington DC 20560, USA.

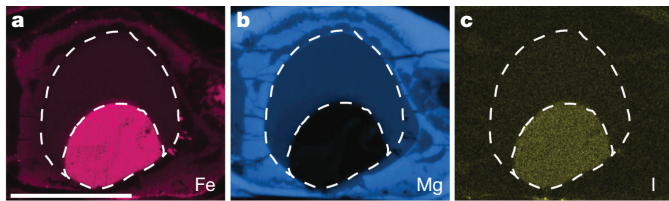


Figure 1 | Compositional maps of Fe, Mg and I from the sample recovered from DAC_I_EXP5, spot 3 at 4,500 K and 40 GPa. a, Fe map; b, Mg map; and c, I map. The smaller dashed outline in each image surrounds the quenched liquid-Fe alloy, the core-analogue material. The larger dashed outline defines the area of the quenched silicate liquid, the magma-ocean-analogue material. The I in c is enriched in core-analogue material and depleted in magma-ocean-analogue material. Maps are taken by energy dispersive spectroscopy (EDS; at 10 kV). Scale bar is 10 μm .

silicate Earth (BSE) values for [FeO] and [W] (Supplementary Table 4), under the endmember scenario that metal segregated from MORB mantle at a temperature along the mantle liquidus²² and occurred as a single event with complete equilibrium between silicate and metal. Major-element chemistry calculations of core–mantle equilibrium follow the approach of ref. 11, and W partitioning is taken from the parameterization reported in ref. 18. Accounting for [W]_{BSE} and [FeO]_{BSE} in a single-stage core formation model requires a Hf/W_{BSE} ratio between 74 and 93 (3.0–3.8 parts per billion (p.p.b.) W_{BSE}), higher than the current best estimate of 18 for primitive MORB source mantle (Methods, Supplementary Table 4).

Given that I is a siderophile element, different episodes of core formation will create early-forming I/Pu variability, and hence $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ heterogeneity in the mantle²⁰ (Fig. 2). A $(2.8 \pm 0.4) \times 1\sigma$ depletion in I for plume mantle would account for the $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ offset between MORB and plume mantle if both reservoirs synchronously closed to Xe loss⁸. A $D^{\text{I}}_{\text{met/sil}}$ value of 14.4 ± 2.1 (1σ) is sufficient to deplete I in plume mantle by $2.8\times$ relative to MORB mantle (the uncertainty on I depletion is not propagated to the $D^{\text{I}}_{\text{met/sil}}$ value) if equal concentrations of I were present during plume and MORB mantle core formation. Larger depletions of I are required to explain the low $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ ratios of plume mantle if plume mantle closed to Xe loss before MORB mantle did. Smaller depletions are needed if I was present at lower abundances during plume mantle core formation. The fact that I is siderophile under high P – T conditions (1) removes the need to interpret low $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ mantle ratios as strictly related to volatile-element depletion during earlier stages of accretion^{2–4} and (2) argues against low W anomalies being related to the incorporation of metal into plume mantle¹⁰ because early-forming, high P – T metal should contain relatively high $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$, opposite to the observations of refs 2–4.

Tungsten becomes less siderophile with increasing P along the mantle liquidus and with increasing oxygen fugacity^{16–18}. Thus, increasing the P for core–mantle equilibration would produce mantle with decreased Hf/W and I/Pu ratios. These parent–daughter fractionations will produce coupled Xe and W anomalies in the mantle with time. We quantify the coupled production of Xe and W anomalies resulting from the partitioning of I and W under a range of P –timing conditions for single-stage core formation (Fig. 3a). Partitioning for I follows from equation (1). Core formation chemistry is calculated in a single-stage framework to limit the number of free parameters and to emphasize the direct effects of partitioning²³. See Methods for additional details regarding the isotopic evolution calculation. The goal here is to identify P –timing conditions for core formation where mantle forms with a $>2.8\times$ depletion in I and with the most extreme W anomalies observed for plume mantle, both higher and lower than MORB mantle^{9,10}.

Mantle that conforms to these geochemical requirements experiences core formation under higher P conditions and earlier compared to MORB mantle (Fig. 3a). High P conditions are required to

Table 1 | Data used to determine the I partitioning parameters in equation (1)

Experiment	Spot number	P/T (GPa K ^{−1})	[S] _{met}	[O] _{met}	[I] _{met}	$\ln(D^{\text{I}}_{\text{met/sil}})$	$\ln(\gamma^{\text{Fe}}_{\text{met}})$
DAC_I_EXP1	1	0.006	0.197	0.038	0.005	−0.15	0.01
DAC_I_EXP1	2	0.0074	0.083	0.119	0.009	1.26	0.25
DAC_I_EXP1	3	0.0073	0.341	0.167	0.019	2.03	1.26
DAC_I_EXP3	1 upper	0.0098	0.063	0.055	0.004	0.76	−0.3
DAC_I_EXP3	1 lower	0.0098	0.065	0.107	0.004	0.85	0.09
DAC_I_EXP3	2	0.0103	0.227	0.158	0.018	1.85	0.62
DAC_I_EXP3	3	0.0088	0.258	0.063	0.002	0.92	0.19
DAC_I_EXP5	2	0.0094	0.224	0.275	0.004	1.8	1.36
DAC_I_EXP5	3	0.0088	0.319	0.181	0.005	2.56	0.93
DAC_I_EXP5	4	0.0082	0.3	0.062	0.01	1.47	0.13
DAC_I_EXP8	1	0.0092	0.145	0.298	0.028	2.35	0.91
DAC_I_EXP8	2	0.01	0.189	0.223	0.024	2.21	0.81
DAC_I_EXP9	4	0.0089	0.001	0.044	0.003	−1.14	−0.1
DAC_I_EXP10	2	0.0092	0.001	0.038	0.008	−0.38	−0.11
DAC_I_EXP10	5	0.0076	0.001	0.04	0.006	−0.76	−0.12
DAC_I_EXP11	2	0.0095	0.103	0.192	0.019	1.55	0.83
DAC_I_EXP11	4	0.0081	0.365	0.168	0.015	2.64	1.2
DAC_I_EXP13	2	0.006	0.35	0.191	0.007	−0.29	1.36

Each data category (for example, P/T) used in equation (1) was identified as significantly correlated with variations in $\ln(D^{\text{I}}_{\text{met/sil}}) + \ln(\gamma^{\text{Fe}}_{\text{met}})$ at the 2σ confidence level using an unweighted, stepwise fitting routine. A complete listing of compositional, pressure, temperature, and uncertainty metadata associated with each DAC heating spot is provided in Supplementary Table 1.

deplete I and evolve to the lower $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ ratios associated with plume mantle^{2–4,8}. Early metal segregation from plume mantle is required to evolve the observed high and low W anomalies^{9,10}. Tungsten isotope evolution in plume mantle follows a two-stage calculation (single-stage core formation). Within the endmember scenario of single, discrete metal segregation events for different mantle reservoirs, the MORB mantle age corresponds to the final step of core formation, and earlier plume mantle ages imply that core formation for these mantle reservoirs occurred before accretion was complete.

Core formation scenarios that successfully account for Xe and W anomalies also produce plume mantle with high FeO contents (Fig. 3). FeO-rich, dense mantle would be robust to mixing with other mantle reservoirs¹⁹, and this may explain why plume mantle preserves isotopic anomalies associated with short-lived decay. High FeO contents are a direct result of the high P – T metal–silicate equilibrium^{11,12} that is required to explain the $>(2.8 \pm 0.4)\times$ depletion in I. A crucial facet to the modelling presented here is that I, a highly volatile element, was present during core formation for both plume and MORB mantle rather than being added only during the late veneer²⁴ (see Methods for additional discussion).

In addition to being FeO-rich, mantle that experienced core formation under high P – T conditions would form with elevated abundances of moderately siderophile elements (MSEs), including W, Ni and Co (Extended Data Fig. 3), and high oxygen fugacity¹¹. Accordingly, this model predicts positive correlations between the magnitude of W anomalies and the abundances of MSEs in plume mantle. Such correlations have not been resolved, and the potential contribution of plume mantle to the bulk silicate Earth MSE budget is highly variable owing to the uncertainty for plume mantle volume and MSE concentrations (see Methods). We emphasize that plume mantle is dominated by Xe recycled from the atmosphere^{2–4}. The presence of this recycled Xe component demonstrates that the geochemistry of plume mantle mostly reflects recycled materials, not accretion processes, and that geochemical correlations associated with accretion will be obscured. The preservation of Xe and W isotopic anomalies may be related to the

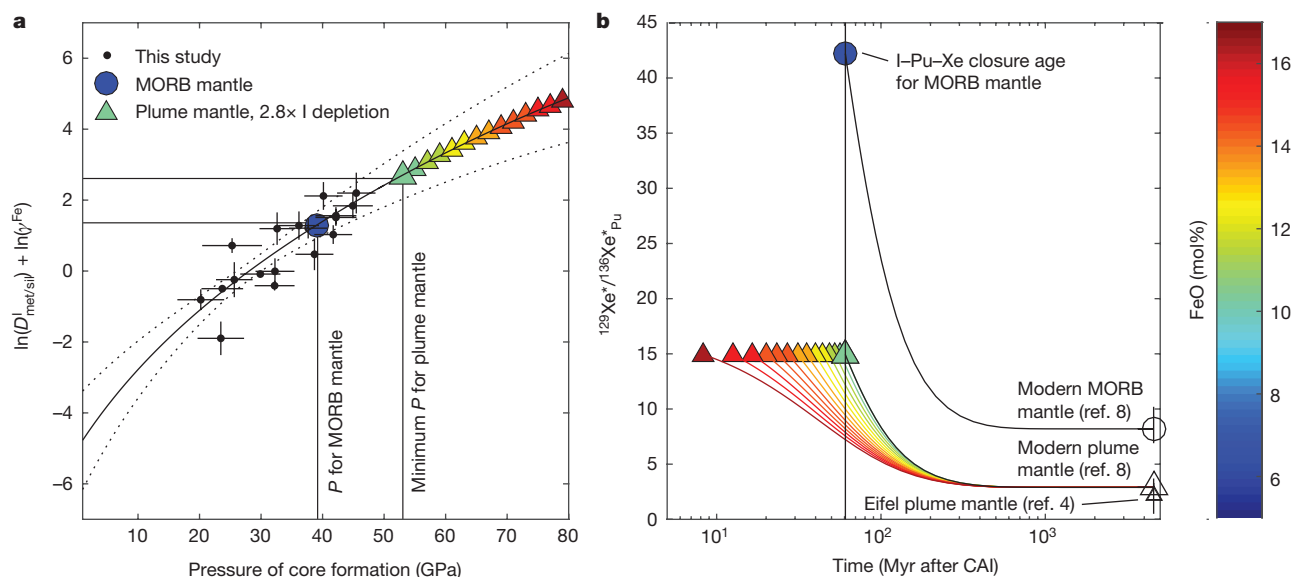


Figure 2 | I partitioning (γ^{Fe} -corrected) during core formation with corresponding Xe isotopic evolution in different mantle domains. **a, Measured $D^I_{\text{met/sil}}$ values (dots, 1σ horizontal and vertical uncertainties) from this study are corrected to the mantle liquidus geotherm²² and O content of the Fe alloy (S-free) in equilibrium with silicate along the liquidus geotherm. Predicted $D^I_{\text{met/sil}}$ values are plotted as the solid line:**

fluid mobility of Xe and W during subduction and their corresponding inefficient recycling^{2,25}.

Linking short-lived isotopic heterogeneity to discrete episodes of core formation implies that Earth's mantle retains a record of the intermediate stages of its growth (Fig. 4). We summarize our model timeline as follows: Earth grew sufficiently large such that I became siderophile under deep-mantle P - T conditions (Fig. 2). Subsections of the mantle then experienced core extraction under relatively high P - T conditions to generate today's I-depleted plume mantle with higher and lower W anomalies (Fig. 3a). The majority of the mantle (MORB mantle) experienced a lower average pressure of metal segregation. Plume and MORB mantle did not homogenize owing to density differences related

to their differing FeO contents (Figs 2 and 3), allowing for the ingrowth and preservation of Xe and W anomalies. Our model departs from those in which P - T progressively increases throughout accretion and the entire mantle homogenizes^{11,26}. Instead, our model produces an initially heterogeneous mantle consistent with dynamical and observational constraints.

Earth was capable of experiencing earlier stages of high P - T core formation based on several lines of reasoning. Accretion models constrained by the W and Pb isotopic composition of MORB mantle suggest that Earth rapidly accreted to nearly its full size within about 50 million years after the initiation of the Solar System²⁷ and was therefore capable of producing high pressures before the Moon-forming

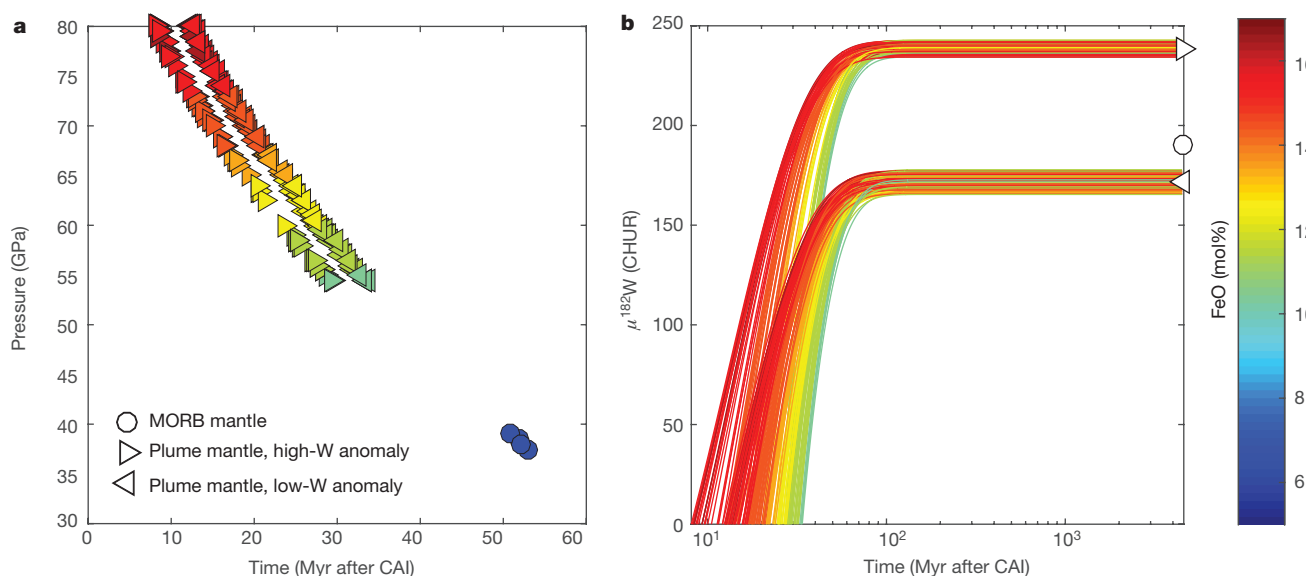


Figure 3 | P -timing conditions of core extraction scenarios that lead to co-evolution of observed Xe and W isotopic anomalies. **a, Specific modelling targets are as follows: (1) a $>2.8\times$ depletion of I relative to MORB mantle and (2) the most extreme observed W isotopic anomalies (right-facing triangles, high anomalies⁹; left-facing triangles, low anomalies¹⁰). The high P - T conditions associated with plume mantle**

formation force the FeO content of plume mantle (colour of symbol; see colour bar) to be correspondingly high. **b**, The modelled W isotopic evolution that results from the P -timing conditions that successfully lead to the observed Xe and W anomalies. CHUR, chondritic uniform reservoir.

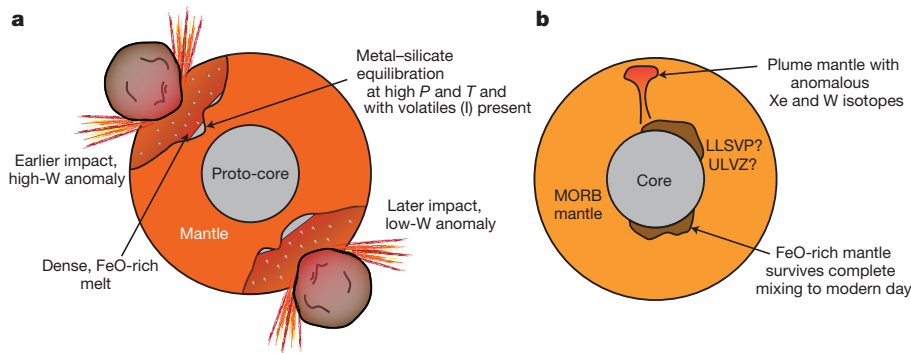


Figure 4 | Cartoon illustrating our model for the co-generation and preservation of Xe and W anomalies. **a**, Earth experiences multiple large impacts; I and other highly volatile elements are present. FeO-rich silicate liquids are produced via high- P , high- T metal-silicate equilibration. The high density of these FeO-rich liquids makes them prone to long-term preservation as distinct geochemical reservoirs that experienced core

impact. Energetic impacts that cause deep magma oceans occur throughout accretion²⁸. Following each impact, metal-silicate equilibrium would be established over the depth range of the impact-induced magma ocean leading to a range of mantle chemistry. In the deepest portions of the magma ocean, metal-silicate equilibration could occur under very high P - T conditions such that the silicate liquid is enriched in FeO and gravitationally stabilized against convective mixing with the rest of the mantle (Fig. 4). The less FeO-rich portions of mantle involved in core formation would be more likely to homogenize and comprise MORB mantle.

High P - T core formation, and the concomitant production of FeO-rich mantle, may also contribute to the formation of large low-shear-velocity provinces and ultralow-velocity zones, which are seismically slow and putatively dense regions of the lowermost mantle usually associated with plume mantle^{13,29,30}. In this context we also note that mantle formed under high P - T conditions, and subsequently stored near the core-mantle boundary, would suffer minimal ^3He degassing. This may explain the high $^3\text{He}/^4\text{He}$ ratios that define high- $^3\text{He}/^4\text{He}$ (plume) mantle^{1,2,14}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 May; accepted 22 November 2017.

- Allège, C. J., Staudacher, T., Sarda, P. & Kurz, M. Constraints on evolution of Earth's mantle from rare gas systematics. *Nature* **303**, 762–766 (1983).
- Mukhopadhyay, S. Early differentiation and volatile accretion recorded in deep-mantle neon and xenon. *Nature* **486**, 101–104 (2012).
- Peto, M. K., Mukhopadhyay, S. & Kelley, K. A. Heterogeneities from the first 100 million years recorded in deep mantle noble gases from the Northern Lau Back-arc Basin. *Earth Planet. Sci. Lett.* **369/370**, 13–23 (2013).
- Caracausi, A., Avive, G., Burnard, P. G., Füre, E. & Marty, B. Chondritic xenon in the Earth's mantle. *Nature* **533**, 82–85 (2016).
- Moreira, M., Kunz, J. & Allegre, C. Rare gas systematics in popping rock: isotopic and elemental compositions in the upper mantle. *Science* **279**, 1178–1181 (1998).
- Holland, G. & Ballentine, C. J. Seawater subduction controls the heavy noble gas composition of the mantle. *Nature* **441**, 186–191 (2006).
- Tucker, J. M., Mukhopadhyay, S. & Schilling, J. G. The heavy noble gas composition of the depleted MORB mantle (DMM) and its implications for the preservation of heterogeneities in the mantle. *Earth Planet. Sci. Lett.* **355/356**, 244–254 (2012).
- Parai, R. & Mukhopadhyay, S. The evolution of MORB and plume mantle volatile budgets: constraints from fission Xe isotopes in Southwest Indian Ridge basalts. *Geochim. Geophys. Geosyst.* **16**, 719–735 (2015).
- Rizo, H. *et al.* Preservation of Earth-forming events in the tungsten isotopic composition of modern flood basalts. *Science* **352**, 809–812 (2016).
- Mundl, A. *et al.* Tungsten-182 heterogeneity in modern ocean island basalts. *Science* **356**, 66–69 (2017).

extraction under distinct P - T - X conditions. We suggest that an earlier impact results in a high-W anomaly, and vice versa, because of the timing offsets depicted in Fig. 3. **b**, Potential modern expression of early-forming reservoirs as components within large low-shear-velocity provinces (LLSVPs) and ultralow-velocity zones (ULVZs), which contain materials with relatively low $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ and variable $^{182}\text{W}/^{184}\text{W}$.

- Rubie, D. C. *et al.* Heterogeneous accretion, composition and core-mantle differentiation of the Earth. *Earth Planet. Sci. Lett.* **301**, 31–42 (2011).
- Frost, D. J., Mann, U., Asahara, Y. & Rubie, D. C. The redox state of the mantle during and just after core formation. *Phil. Trans. R. Soc. Lond.* **366**, 4315–4337 (2008).
- French, S. W. & Romanowicz, B. Broad plumes rooted at the base of the Earth's mantle beneath major hotspots. *Nature* **525**, 95–99 (2015).
- Kurz, M. D., Jenkins, W. J. & Hart, S. R. Helium isotopic systematics of oceanic islands and mantle heterogeneity. *Nature* **297**, 43–47 (1982).
- Kleine, T. *et al.* Hf-W chronology of the accretion and early evolution of asteroids and terrestrial planets. *Geochim. Cosmochim. Acta* **73**, 5150–5188 (2009).
- Walter, M. J. & Thibault, Y. Partitioning of tungsten and molybdenum between metallic liquid and silicate melt. *Science* **270**, 1186–1189 (1995).
- Cottrell, E., Walter, M. J. & Walker, D. Metal-silicate partitioning of tungsten at high pressure and temperature: evidence for equilibrium core formation in Earth. *Earth Planet. Sci. Lett.* **281**, 275–287 (2009).
- Wade, J., Wood, B. J. & Tuff, J. Metal-silicate partitioning of Mo and W at high pressures and temperatures: evidence for late accretion of sulphur to the Earth. *Geochim. Cosmochim. Acta* **85**, 58–74 (2012).
- Sleep, N. H. Gradual entrainment of a chemical layer at the base of the mantle by overlying convection. *Geophys. J. Int.* **95**, 437–447 (1988).
- Armstrong, R. M., Jephcoat, A. P., Bouhifd, M. A. & Porcelli, D. Metal-silicate partitioning of iodine at high pressures and temperatures: implications for the Earth's core and ^{129}Xe budgets. *Earth Planet. Sci. Lett.* **373**, 140–149 (2013).
- Ma, Z. Thermodynamic description for concentrated metallic solutions using interaction parameters. *Metall. Mater. Trans. B* **32**, 87–103 (2001).
- Fiquet, G. *et al.* Melting of peridotite to 140 gigapascals. *Science* **329**, 1516–1518 (2010).
- Walter, M. J. & Cottrell, E. Assessing uncertainty in geochemical models for core formation in Earth. *Earth Planet. Sci. Lett.* **365**, 165–176 (2013).
- Wang, Z. & Becker, H. Ratios of S, Se and Te in the silicate Earth require a volatile-rich late veneer. *Nature* **499**, 328–331 (2013).
- Jackson, M., Shirey, S., Hauri, E., Kurz, M. & Rizo, H. Peridotite xenoliths from the Polynesian Austral and Samoa hotspots: implications for the destruction of ancient ^{187}Os and ^{142}Nd isotopic domains and the preservation of Hadean ^{129}Xe in the modern convecting mantle. *Geochim. Cosmochim. Acta* **185**, 21–43 (2016).
- Wood, B. J., Walter, M. J. & Wade, J. Accretion of the Earth and segregation of its core. *Nature* **441**, 825–833 (2006).
- Rudge, J. F., Kleine, T. & Bourdon, B. Broad bounds on Earth's accretion and core formation constrained by geochemical models. *Nat. Geosci.* **3**, 439–443 (2010).
- Chambers, J. E. & Wetherill, G. W. Making the terrestrial planets: N -body integrations of planetary embryos in three dimensions. *Icarus* **136**, 304–327 (1998).
- McNamara, A. K. & Zhong, S. Thermochemical structures beneath Africa and the Pacific Ocean. *Nature* **437**, 1136–1139 (2005).
- Jackson, M., Konter, J. & Becker, T. Primordial helium entrained by the hottest mantle plumes. *Nature* **542**, 340–343 (2017).

Supplementary Information is available in the online version of the paper.

Acknowledgements Portions of this work were performed at GeoSoilEnviroCARS (The University of Chicago, Sector 13), the Advanced Photon Source, Argonne National Laboratory. GeoSoilEnviroCARS is supported by the National Science Foundation—Earth Sciences (EAR-1128799) and the Department of Energy—GeoSciences (DE-FG02-94ER14466). This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by

Argonne National Laboratory under contract number DE-AC02-06CH11357. We acknowledge the support provided by C. Prescher, E. Greenberg, and V. Prakapenka during the experiments conducted at GSECARS and in data reduction. We acknowledge E. Bullock and S. Vitale for their assistance with the electron microscopy, electron microprobe analysis, and focused ion beam recovery. We thank T. Gooding for his assistance in preparing the piston cylinder experiments for analysis. We acknowledge discussions with R. Fischer, R. Parai, S. Parman, J. Tucker and M. Nakajima during the development of the manuscript. C.R.M.J., N.R.B. and Z.D. acknowledge fellowship support from the Carnegie Institution for Science, and C.R.M.J. additionally acknowledges fellowship support from the Smithsonian Institution. The research was also supported by NSF grants to C.R.M.J. (EAR-1725315), to E.C. (EAR-0738654) and to Y.F. (EAR-1447311).

Author Contributions C.R.M.J. conceived the project. C.R.M.J. conducted the experiments and analyses in collaboration with N.R.B. and Z.D. C.R.M.J., N.R.B.

and Z.D. all contributed to the model framework. N.R.B. conducted the W isotope and major-element chemistry calculations. C.R.M.J., N.R.B., Z.D., E.C. and Y.F. contributed to data interpretation, presentation and manuscript preparation. E.C. facilitated the project and piston cylinder experiments. Y.F. facilitated the project and DAC experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to C.R.M.J. (jacksonc3@si.edu).

Reviewer Information *Nature* thanks G. Avice and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

DAC experiments. Preparation. All DAC experiments used pre-indented Re gaskets. Gaskets for experiments under 25 GPa were indented to 35 μm thickness, and gaskets for experiments above 25 GPa were indented to 20 μm thickness. A 130- μm sample chamber was cut into each gasket using a laser drill following indentation. A synthetic silicate–metal mixture with a composition of C1/C^{31} was ball-milled and then mixed with 5% $\text{KI} \pm 20\%$ FeS using a mortar and pestle. Given the small size of each laser heating spot, the fact that the entire sample is not melted, and the ubiquitous presence of C, the bulk composition of each laser heating spot is highly variable. These starting materials were loaded into sample chambers and then compressed using stepped anvils machined to the diameter of the sample chamber³². The height of the stepped anvils is about 10 μm . The remaining portions of the sample chamber were filled using pressed foils of MgO. This creates a three-layer arrangement with two layers of MgO about 10 μm thick surrounding a silicate–metal mixture approximately 10 μm thick. Gaskets were then loaded into a DAC and placed in a vacuum oven for overnight heating at 110 °C to minimize adsorbed water on sample powders. Immediately after removal from the vacuum oven, gaskets were compressed to moderate pressure. Once the DAC had cooled to room temperature, the gaskets were compressed to the desired experimental pressure. Pressure upon compression was monitored using the diamond edge technique³³. **Laser heating and X-ray diffraction (XRD).** DAC sample heating was completed at beamline 13-ID-D of the GSECARS facility at the Advanced Photon Source using a double-sided, flat-top laser heating system ($\lambda = 1,064 \text{ nm}$) with a diameter of about 20 μm . Emission spectra were collected for both sides of the heating spot using light from a 2 $\mu\text{m} \times 2 \mu\text{m}$ region co-aligned with the XRD measurements. Temperatures were calculated from the emission spectra, assuming greybody behaviour. Temperatures reported here used light collected between 670 nm and 770 nm. Including longer-wavelength data reduced the quality of fit and generally resulted in lower temperatures (about 200 K). Each reported temperature is the average of a series of measurements taken in rapid succession. At high temperature (>3,000 K), five spectra are taken for each series per side of the heating spot. Temperatures reported for an experiment are the average of the final series of temperature measurements on the hotter of either the upstream or downstream sides. We take the hotter of the two temperatures as there is the potential for small misalignments between the hottest portion of the heating spot and the region where light was collected for the spectral measurement. Any misalignment creates bias towards cooler temperatures. Uncertainties in temperature are the standard deviations (1σ) of the final series of temperature measurements on the hotter side of the DAC plus an analytical uncertainty of 100 K. Heating cycles generally lasted several minutes and heating spots were held at peak or near-peak temperatures for >10 s. Temperatures for DAC experiments are reported in Supplementary Table 1.

XRD data were collected for every heating spot before heating and with every temperature measurement during heating. These data we processed with Dioptas software³⁴ to calculate pressure before heating ('cold pressure') and pressure during the heating cycle (Supplementary Table 1). We use the unit cell volume of MgO, which was present as the pressure medium in all experiments, to calculate the cold pressure.

The pressure during the heating cycle deviates from the cold pressure owing to relaxation of the sample and thermal expansivity. The combination of these effects generally results in a 10–20% increase in pressure upon heating to melting temperatures at mid-mantle pressures³⁵.

At high temperatures (>3,000 K) in the experiments reported here, a second set of MgO peaks are present within the XRD spectra. This second set of MgO peaks are shifted towards larger volume compared to the MgO peaks of the pressure medium. We interpret the larger-volume MgO peaks to be the Fe-bearing MgO phase that mantles all heating spots in the present experiments (Fig. 1), and these peaks are used to calculate thermal pressure. To calculate pressure, we apply the volume measurements from the XRD data and the temperature measurements to the thermal equation of state for MgO from ref. 36. This approach neglects the FeO present in the mantling MgO phase (magnesium number Mg\# of about 80) and the temperature gradient across the mantling phase. These two effects cancel to a degree. This approach has the advantage of being able to monitor thermal pressure over the duration of the experiment. Uncertainty on the pressure measurement is the fractional uncertainty of the temperature measurement multiplied by the hot pressure plus 2 GPa (Supplementary Table 1).

Heating spots DAC_I_EXP5 spot 4, DAC_I_EXP11 spot 4, and DAC_I_EXP13 spot 2 all have thermal pressure in excess of 30% initial pressure. This is larger than normally attributed to thermal pressure in core formation DAC experiments³⁷. We suggest that the large thermal pressure relates to the pre-relaxation of the heating spot due to previous heating spots completed in the same cell.

Chemical analysis preparation. The heating spots of the DAC experiments were prepared for analysis using a focused ion beam (30 kV, Ga^+ ions, Auriga, Zeiss Instruments) at Carnegie Institution for Science, Washington DC following the

technique of ref. 38. Samples were milled to expose a cross-section of the heating spot parallel to the laser beam used to heat the sample. Milling proceeded on each heating spot until quenched metal and quenched silicate phases were exposed and the exposure of the metal phase had nearly reached its full diameter. The final milling step was completed with a 2-nA beam to prepare a flat surface for wavelength dispersive spectroscopy analysis and EDS mapping.

Piston cylinder experiments. Preparation. A series of piston cylinder experiments were conducted at the Department of Mineral Sciences, National Museum of Natural History, USA, to corroborate the DAC results. Experiments were run using 0.5-inch-diameter assemblies, comprising BaCO_3 pressure media wrapped in Pb foil, straight-walled graphite heaters, and MgO spacers. Temperatures were monitored using a D-type thermocouple housed in a four-hole alumina rod inserted axially into the assembly. MgO spacers were machined to position the hotspot between the sample capsule and the thermocouple junction. Starting materials were oxide and carbonate powders mixed to generate a basaltic composition (62A from ref. 39). This powder was decarbonated and reduced to the oxygen fugacity of mostly iron–wüstite (IW) at 1,373 K overnight. Following decarbonation and reduction, the oxide was then mixed with either FeNiSi alloy, FeSi alloy, $\text{FeSi} + \text{FeS}$ or $\text{Fe} + \text{FeS}$ to yield a 50:50 silicate–metal mixture by weight. I was added as KI, AgI or I to yield approximately 4 wt% I in the starting material. Graphite capsules were used in all experiments. Capsules were isolated from the graphite heater by an Al_2O_3 sheath, and the thermocouple was isolated from the graphite capsule by a 1-mm-thick disk of MgO.

Samples were cold-compressed to 1 GPa. Following compression, sample temperatures were ramped to 1,073 K and allowed to sinter for 2 h. Experiments were then ramped to 1,973 K and held at that temperature for 20–40 min. Experimental pressures were 1.5 GPa.

Chemical analysis preparation. Piston cylinder experiments were prepared to minimize the loss of fluid-mobile components from the run products. A basaltic silicate composition was chosen to enable the silicate to quench to a glass. Graphite capsules were sectioned using a diamond wafering blade without lubrication. Initial exposure of run products was completed with SiC sandpaper using Turbinoil as a lubricant. Final polishing was completed using alumina-impregnated sheets without lubricant.

A series of piston cylinder experiments were run and prepared with water-based lubrication. In these experiments, I-rich materials were observed on the polished surfaces following the final polishing step. We do not report these experiments because of the clear mobility of I before analysis.

Chemical analysis. Chemical analyses were completed using a JEOL 8530F field-emission microprobe at the Carnegie Institution for Science. Analyses of metal phases were performed with a beam diameter ranging from a focused beam to a 2- μm -diameter beam. Analyses on silicate materials were conducted with a 1- μm -diameter beam. All analyses employed a 10-kV accelerating voltage and 2–10-nA beam current. Repeated analyses of the same material with different current densities yielded no evidence for the volatilization of I during microanalysis. Analytical standards for silicate analyses were ENAL5 (enstatite stoichiometry glass with 5 wt% Al_2O_3 , for Mg, Si and Al), basalt 812 glass (for Ca, K, Fe), FeS_2 (for S), and KI (for I). Analytical standards for metal analyses were ENAL5 (for Mg, Si and Al), basalt 812 glass (for Ca and K), FeS_2 (for S), KI (for I), Fe_7C_3 (for C and Fe), and Fe_3O_4 (for O). The Fe_7C_3 standard was synthesized following the method outlined in ref. 40. Multiple analyses were conducted on the silicate phase and metal phase contained within each heating spot. Analytical uncertainties are calculated as standard errors of the repeat analyses of a given material. Standard errors are reported because the quench textures of metals and silicate can be coarse with respect to the beam diameter. Heating spots with small metallic phases were analysed and repolished with the focused ion beam to allow for repeat analyses. Reanalysis of metals without polishing resulted in analyses with carbon concentrations about 1 wt% higher compared to the original analysis.

The glass phases of the piston cylinder experiments were analysed using the same routine developed for the silicate phases in the DAC. The metallic phases of the piston cylinder experiments were analysed with a 15-kV accelerating voltage and 50-nA beam current, with the beam defocused to 10 μm , to achieve a lower detection limit for I. Standards were the same as the 10-kV routine with the following additions: Ag metal (Ag), basalt 812 glass (Na and Ti), Ni-olivine (Ni), MnO_2 (Mn) and chromite (Cr). All samples and standards were coated with Ir to prevent charging and facilitate the quantification of C in metal.

The compositions of DAC experimental products are reported in Supplementary Table 1, and the compositions of piston cylinder experimental products are reported in Supplementary Table 2.

Data processing and fitting. To parameterize $D_{\text{met/sil}}^{\text{I}}$ values we consider the dissolution partitioning reaction: $[\text{I}]_{\text{sil}} = [\text{I}]_{\text{met}}$. We take this partitioning reaction as opposed to other potential reactions because the speciation of I dissolved in liquid Fe alloy and silicate is not known. The partitioning reaction has an associated equilibrium constant, K_{I} .

$$K_I = \gamma_{\text{met}}^I [I]_{\text{met}} / \gamma_{\text{sil}}^I [I]_{\text{met}} \quad (3)$$

$$-K_I = \exp \left(\frac{\Delta H^0}{RT} - \frac{\Delta S^0}{R} + \frac{P\Delta V^0}{RT} \right) \quad (4)$$

where γ denotes the activity coefficient, and $[]$ denotes the concentration of a given component. ΔH^0 , ΔS^0 and ΔV^0 are the enthalpy, entropy and volume change of the partitioning reaction for a given pressure and temperature. Assuming ΔH^0 , ΔS^0 and ΔV^0 are constant across the pressure and temperature range of interest and that γ_{sil}^I is a constant, the I partition coefficient ($D_{\text{met/sil}}^I = [I]_{\text{met}}/[I]_{\text{sil}}$) can be expressed as follows:

$$\ln(D_{\text{met/sil}}^I) + \ln(\gamma_{\text{met}}^I) = a + \frac{b}{T} + \frac{cP}{T} \quad (5)$$

where a , b and c are fitting parameters that scale with ΔS^0 , ΔH^0 and ΔV^0 , respectively. Following from equation (24) in ref. 21, γ_{met}^I is expanded as:

$$\ln(\gamma_{\text{met}}^I) = \ln(\gamma_{\text{met}}^{\text{Fe}}) + \ln(\gamma_{\text{met}}^{\text{I}_0}) - \sum \varepsilon_i^I X_{\text{met}}^i \quad (6)$$

where $\gamma_{\text{met}}^{\text{I}_0}$ denotes the activity coefficient of I at infinite dilution and ε are the fitted interaction parameters. The sigma terms sum across the alloying components in Fe alloy and X_{met}^i takes the form:

$$X_{\text{met}}^i = T_r/T \left[[i]_{\text{met}} \left(1 + \frac{\ln(1 - [i]_{\text{met}})}{[i]_{\text{met}}} - \frac{1}{1 - [I]_{\text{met}}} \right) - [i]_{\text{met}}^2 [I]_{\text{met}} \left(\frac{1}{1 - [I]_{\text{met}}} + \frac{1}{1 - [i]_{\text{met}}} + \frac{[I]_{\text{met}}}{2(1 - [I]_{\text{met}})^2} - 1 \right) \right] \quad (7)$$

for $i = \text{S, C, O}$ or Re and the reference temperature $T_r = 1,873 \text{ K}$.

$$X_{\text{met}}^I = (T_r/T) \ln(1 - [I]_{\text{met}}) \quad (8)$$

Known terms are collected on the left and unknown terms are collected on the right, to yield the final expression used for parameterization:

$$\ln(D_{\text{met/sil}}^I) + \ln(\gamma_{\text{met}}^{\text{Fe}}) = a + \frac{b}{T} + \frac{cP}{T} + \sum \varepsilon_i^I X_{\text{met}}^i \quad (9)$$

Variations in $\ln(D_{\text{met/sil}}^I) + \ln(\gamma_{\text{met}}^{\text{Fe}})$ were fitted in a stepwise, unweighted least-squares approach. The stepwise matrix includes the $1/T$, P/T , $X_{\text{met}}^{\text{S}}$, $X_{\text{met}}^{\text{C}}$, $X_{\text{met}}^{\text{O}}$ and $X_{\text{met}}^{\text{Re}}$ terms. The a term accounts for $\Delta S^0/R$ and is forced to be part of the parameterization.

In the first fitting step, the most significant term is $X_{\text{met}}^{\text{O}}$ (with a P value of 0.00005, Extended Data Fig. 1a). Following the addition of $X_{\text{met}}^{\text{C}}$, the most significant term is $X_{\text{met}}^{\text{S}}$ (with a P value of 0.02), and following the addition of $X_{\text{met}}^{\text{Re}}$, the most significant term is P/T (with a P value of 0.007). No additional terms are significant at the 95% confidence threshold and are therefore not included in the predictive expression for I partitioning (equation (1)). Each step of the stepwise regression is plotted in Extended Data Fig. 1. The fact that X_{met}^I is not significant in the stepwise fitting approach suggests that I doping levels are not high enough to significantly affect $\ln(D_{\text{met/sil}}^I)$ values; that is, doping levels are within the Henrian regime. The activity coefficient of I at infinite dilution (equation (5), $\gamma_{\text{met}}^{\text{I}_0}$) in liquid Fe metal is not explicitly accounted for in calculating the solution behaviour of I . This term has the same $1/T$ dependence as the b term, which is not deemed significant and is therefore not included in the current parameterization. The parameter values and the associated covariance matrix are included as Supplementary Table 3. Covariance plots of parameterization terms are provided in Extended Data Fig. 1.

All silicate and metal data were converted to atomic fractions to calculate $D_{\text{met/sil}}^I$. Oxygen was subtracted from the silicate analyses according to the valence state of the anionic species, S^{2-} and I^- .

Atomic data were converted to molar oxide for the silicate phase of each experiment to calculate oxygen fugacity and exchange coefficients. The substitution of S and I for O is not accounted for in this conversion. The atomic fraction of S relative to Fe in experiments is approximately 0.1. Correspondingly, any reduction to the mole fraction of FeO that is due to Fe complexing with S is small.

Oxygen fugacity for each experiment was calculated with respect to the iron-wüstite oxygen buffer as:

$$\Delta \text{IW} = 2 \log_{10} (X_{\text{sil}}^{\text{FeO}} \gamma_{\text{FeO}}^{\text{Fe}} / X_{\text{met}}^{\text{Fe}} \gamma_{\text{met}}^{\text{Fe}}) \quad (10)$$

We take the expression from ref. 41 for $\ln(\gamma_{\text{sil}}^{\text{FeO}})$:

$$\ln(\gamma_{\text{sil}}^{\text{FeO}}) = 2,096/T - 2.6024X_{\text{sil}}^{\text{FeO}} + 2.2105X_{\text{sil}}^{\text{SiO}_2} + 0.238X_{\text{sil}}^{\text{CaO}} - 0.9666 \quad (11)$$

Calculating $\ln(\gamma_{\text{met}}^{\text{Fe}})$ requires an understanding of how different components interact in liquid Fe alloy, and we follow the formalism for calculating $\gamma_{\text{met}}^{\text{Fe}}$ outlined in ref. 21. Non-Fe components in the alloy are limited to O , S and C . Interactions are calculated using the ε values tabulated in the Steelmaking Data Sourcebook⁴². We do not account for I , Re , Mg , K and Al interactions in calculating $\gamma_{\text{met}}^{\text{Fe}}$. These elements are present at minor to trace levels, except for Re , which is present at an atomic fraction of about 0.15 in the metallic phases of three heating spots. Values of $\gamma_{\text{met}}^{\text{Fe}}$ were also calculated using the METALACT calculator (MetalAct, <http://www.earth.ox.ac.uk/~expet/metalact/>; ref. 43) with Si , S , O , C and Re interactions. There is good agreement between the two sets of calculations (not shown).

Low totals in the silicate of DAC heating spots. Several heating spots yielded low totals during analysis of the silicate (DAC_I_EXP5 spot 2, DAC_I_EXP5 spot 4 and DAC_I_EXP11 spot 4; Supplementary Table 1). Following the analysis of DAC_I_EXP5 spot 2 and DAC_I_EXP5 spot 4, vesiculation of the silicate was observed (Extended Data Fig. 4). We interpret this to indicate that the quenched silicate phases in these heating spots were volatile-rich. Low totals during analysis of DAC_I_EXP11 spot 4 may relate to sample specific geometry issues. Ferropericlase and bridgmanite mineral grains were analysed in the region immediately surrounding heating spots with low totals (DAC_I_EXP5 spot 2, DAC_I_EXP5 spot 4 and DAC_I_EXP11 spot 4; not reported). These analyses consistently yielded higher totals than the quenched silicate. This provides additional evidence that the low totals may be related to unanalysed volatiles and are not likely to be related to geometric issues or analytical routine issues.

We note that changes to the major element compositions of ultramafic melts do not appear to have substantial effects on metal-silicate partition coefficients. Following other studies⁴³, we assume ultramafic melt composition does not affect metal-silicate partition coefficients. The effects of C and H dissolved in ultramafic melts have not been quantified, but we assume from the results with major elements that C and H do not affect partitioning behaviour.

Evaluating equilibrium in DAC experiments. The approach to equilibrium in the present experiments is evaluated by (1) comparing the exchange systematics of the present experiments to previously reported data and (2) reviewing the internal systematics of the present experiments.

Extended Data Fig. 5a compares measurements of Si-Fe exchange coefficients between silicate and metal ($K_{\text{D}}^{\text{Si-Fe}}$) from the present experiments to $K_{\text{D}}^{\text{Si-Fe}}$ values compiled by ref. 38. We choose $K_{\text{D}}^{\text{Si-Fe}}$ for comparison because it has a high and well-defined temperature dependence and because the compositional effects are relatively small and well constrained³⁷. Data from this study have been corrected for Si-Si , Si-O , Si-C and Si-S interactions, and data from the literature have been corrected for Si-Si , Si-O and Si-C interactions. Corrections for Si-Si , Si-C and Si-S were made using the ε parameters reported in ref. 42. Corrections for Si-O were made using the ε parameter reported by ref. 37. Reported and literature $K_{\text{D}}^{\text{Si-Fe}}$ values show good correspondence (Extended Data Fig. 5a), suggesting that we heated our samples stably for long enough to closely approach chemical equilibrium and to obtain accurate temperature calculations.

The Mg/Si ratio of eutectic melts in the MgO-MgSiO_3 system increases with pressure up to 45 GPa (ref. 44; Extended Data Fig. 5b). This behaviour provides a check on pressure estimates for individual heating spots. Circle symbols are from this study and are binned into three groups by Mg\# . Darker symbols denote a lower Mg\# . Silicate Mg\# spans from 61 to 77. Square symbols have a Mg\# of 100 and are from ref. 44.

Compared to the silicate compositions of MgO-MgSiO_3 eutectic melts reported by ref. 44 at 35 GPa and 45 GPa (Extended Data Fig. 5b), the silicate liquids from this study have uniformly higher $(\text{Mg} + \text{Fe})/\text{Si}$ ratios at similar pressures. More FeO -rich silicate liquids within a given pressure range also plot towards higher $(\text{Mg} + \text{Fe})/\text{Si}$ ratios. These behaviours are expected for systems of similar pressure but variable FeO content. The addition of FeO to a system lowers the melting point of MgO relative to MgSiO_3 , shifting the eutectic towards MgO . Further support for the pressure calculation comes from the similar slopes of $(\text{Mg} + \text{Fe})/\text{Si}$ versus pressure for the Mg\# groups (symbols of same colour) and the eutectic melt slope⁴⁴.

Mobility of I during sample preparation. Experiments conducted with a DAC were prepared for analysis exclusively using a focused ion beam. No contact was made between the samples and a fluid that may promote I mobilization. Nonetheless, many DAC samples developed I -rich blebs in the vicinity of the exposed heating spots during storage in desiccators (Extended Data Fig. 6). Heating spots were analysed using EDS before their removal from the focused ion beam (before the appearance of I -rich blebs). The I contents of the metals and silicates before and after the appearance of I -rich blebs are indistinguishable. This indicates

that the I-rich blebs were not derived from the silicate or metal phases and that the I-rich blebs did not modify the I content of the silicate and metal phases.

EDS analyses of the I-rich blebs indicate that they are rich in Fe, C, I and O. In DAC experiments without S, a material with a similar composition mantles the metal phases (Extended Data Fig. 6) and can extend into the region immediately surrounding the heating spot. We consider this material to be a phase that is separate to the metal and silicate phases. If this I-rich material exsolved from the metal or silicate during quenching in the S-free experiments (the mantling phase is not present in S-bearing experiments), this would imply a very O-rich and I-rich metal or silicate. This possibility is not supported by the partitioning behaviour of O and I in S-bearing experiments. Specifically, I partitioning for S-free experiments is well predicted by S-bearing experiments, and O solubility in metals from S-free experiments is similar to those measured in S-bearing experiments. Moreover, the fact that I-rich blebs are not always observed in direct contact with the silicate or metal phase of a heating spot suggests an independent origin for the I-rich material (Extended Data Fig. 6).

Comparison of DAC and piston cylinder results. A series of piston cylinder experiments were completed to validate the $\ln(D_{\text{met/sil}}^I)$ expression derived using the DAC results (equation (1)). All experiments were conducted at 1.5 GPa and 1,973 K. In all but two experiments the concentration of I in the metal phase was below the detection limit, precluding the quantification of $\ln(D_{\text{met/sil}}^I)$ values. For the analytical parameters (15 kV, 50 nA, 60 s/30 s peak/off-peak counting time) used to measure I in metals of the piston cylinder experiments, the detection limit is 0.006 wt% or 60 parts per million. We take the detection limit to calculate upper limits on $\ln(D_{\text{met/sil}}^I)$ values for these experiments (Extended Data Fig. 2).

I partition coefficients were quantified for two S-bearing piston cylinder experiments. The S-bearing experiments contained two, immiscible metallic liquids; one S-rich and the other S-poor. I is detectable only in the S-rich metallic phase, consistent with S in metallic liquids increasing $\ln(D_{\text{met/sil}}^I)$ values (equation (1)). The two S-bearing experiments were conducted at $\Delta IW = 4.7$ and $\Delta IW = 1.9$, where $\Delta IW + a$ is fugacity in log units relative to the iron–wüstite oxygen buffer. The more reduced experiment yielded a $\ln(D_{\text{met/sil}}^I)$ value that is lower than the predicted range for $\ln(D_{\text{met/sil}}^I)$ at low pressure (95% confidence threshold, Extended Data Fig. 2a). The more oxidized experiment yielded a higher $\ln(D_{\text{met/sil}}^I)$ value that is within the predicted $\ln(D_{\text{met/sil}}^I)$ range. Oxygen fugacity (f_{O_2}) in the DAC experiments averaged $\Delta IW = 1.3$, similar to the more oxidized piston cylinder experiment ($\Delta IW = 1.9$). Thus, the piston cylinder experiments provide independent support for the predictive ability of the DAC regression for systems with similar f_{O_2} . The fact that the measured $\ln(D_{\text{met/sil}}^I)$ values and upper limit $\ln(D_{\text{met/sil}}^I)$ values plot mostly near the lower limit of the uncertainty envelope favours models with larger P/T terms and lower intercepts (Extended Data Fig. 2b).

From the two piston cylinder experiments, there is preliminary support for oxidizing conditions promoting I partitioning into metallic liquids. This behaviour is consistent with that of other anionic species in core-forming environments^{45–48}. Taking the two piston cylinder experiments with measurable I in the metallic phase as a guide for how $\ln(D_{\text{met/sil}}^I)$ scales with f_{O_2} allows the piston cylinder data to be corrected to the average f_{O_2} of the DAC experiments ($\Delta IW = 1.3$). The f_{O_2} -corrected piston cylinder data are plotted in Extended Data Fig. 2b.

We choose not to include an f_{O_2} term in the DAC data regression because (1) we apply equation (1) under f_{O_2} similar to the DAC data, (2) the f_{O_2} term is not statistically significant (P value > 0.05) within the stepwise fit of the DAC data, and (3) a systematic study quantifying the relationship between f_{O_2} and $\ln(D_{\text{met/sil}}^I)$ has not been completed.

Modelling Xe and W isotopic evolution. Establishing the I/Pu fractionations required to produce $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ variability between MORB and plume mantle. The pertinent isotopic observations comparing Xe isotopes in MORB and plume mantle are as follows: MORB and plume mantle $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ values are $8.2^{+2}_{-1.3}$ and $2.9^{+0.4}_{-0.1}$, respectively⁸, that is, the $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ ratio for MORB mantle is $2.8 \pm 0.4 \times$ higher than in plume mantle. We note the uncertainties in MORB and plume mantle $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ values are similarly asymmetric such that their ratio has symmetric uncertainty at the precision of the measurements. Isotopic variability within $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ reflects a combination of different Xe closure ages for plume and MORB mantle and the I/Pu ratios of these two reservoirs during the lifetime of the parent isotope. The Xe closure age (I–Pu–Xe) is the age at which a reservoir starts to accumulate radiogenic Xe related to I and Pu decay. Closure age is calculated as:

$$t_{129/244}^{\text{closure}} = (1/\lambda_{244} - \lambda_{129}) \ln \left(\frac{^{129}\text{Xe}^*}{^{136}\text{Xe}^*_{\text{Pu}}} \times \frac{^{238}\text{U}_i}{^{127}\text{I}_i} \times \frac{^{244}\text{Pu}_i}{^{238}\text{U}_i} \div \frac{^{129}\text{I}_i}{^{127}\text{I}_i} \times Y_{244} \right) \quad (12)$$

where $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ is determined for different geochemical reservoirs using average carbonaceous chondrite (AVCC) values for initial mantle Xe (taken from

ref. 8) and the remaining parameters are reported in ref. 49. We use the AVCC-based calculations because AVCC provides the best fit⁴ to the non-radiogenic components to mantle Xe. The ^{127}I contents of MORB mantle are updated to reflect the most recent estimates⁵⁰ of 7 ± 3 p.p.b. I for the bulk silicate Earth.

If MORB and plume mantle closed to Xe loss with the same I/Pu ratios, then MORB mantle must have closed to Xe loss about 30 million years earlier than plume mantle to account for its $2.8 \times$ higher ratio of $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ (Extended Data Fig. 7a). The apparent common heritage of the W isotopes of MORB mantle and the Moon^{51,52}, however, suggests that MORB mantle closed to Xe loss after the plume mantle did. Lowering the I/Pu ratio of plume mantle progressively lowers its Xe closure age. Equal Xe closure ages for plume and MORB mantle are accounted for if the plume mantle closed with a $2.8 \times$ lower I/Pu ratio compared to MORB mantle (Fig. 2b). Greater depletions of I/Pu for plume mantle translate to Xe closure ages that predate Xe closure in MORB mantle (Fig. 2b).

We note that the radiogenic component of the $^{129}\text{Xe}/^{130}\text{Xe}$ ratio for plume mantle, corrected for recycling, is also approximately $3 \times$ lower than MORB mantle (Extended Data Fig. 7b, Supplementary Table 4). The radiogenic component of the $^{129}\text{Xe}/^{130}\text{Xe}$ ratio is calculated as follows:

$$^{129}\text{Xe}/^{130}\text{Xe}_{\text{radiogenic}} = ^{129}\text{Xe}/^{130}\text{Xe}_{\text{mantle-recycling-corrected}} - ^{129}\text{Xe}/^{130}\text{Xe}_{\text{initial}} \quad (13)$$

where $^{129}\text{Xe}/^{130}\text{Xe}_{\text{initial}} = 6.286$ (ref. 49). The similar offsets for $^{129}\text{Xe}/^{130}\text{Xe}_{\text{radiogenic}}$ and $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ between plume and MORB mantle imply that they are related to the behaviour of I, and not Pu or Xe, during accretion.

Developing W isotopic targets. Xenon from MORB mantle and plume mantle is dominantly from the atmosphere^{2–4,6–8}. The injection of atmospheric Xe into the mantle occurs during subduction of slab materials that have interacted with oceanic waters. These materials contain W from the MORB mantle, and therefore, the injection of atmospheric Xe into the mantle also contains a component of MORB W. The MORB mantle $^{182}\text{W}/^{184}\text{W}$ ratio is not affected by this process, but backmixing MORB W into plume mantle will dilute any $^{182}\text{W}/^{184}\text{W}$ anomaly present in plume mantle and change its concentration of W.

The degree that plume $^{182}\text{W}/^{184}\text{W}$ anomalies are diluted owing to backmixing of MORB W depends on the coupling between Xe and W during their respective deep cycles. Both W and Xe are efficiently stripped from the slab because they are incompatible in mantle and slab minerals and fluid that is mobile during slab subduction^{53–55}. Alteration of oceanic crust results in a net uptake of Xe into the slab before subduction, increasing Xe concentrations in oceanic crust by $100 \times$ to $100,000 \times$ over Xe concentrations in MORB mantle^{56–60}.

The W/Th ratio for arc lavas affected by slab melts is essentially equal to the W/Th ratio for MORB⁶¹. Given that W and Th are incompatible during slab melting, the equality of W/Th in the slab melt and in MORB suggests minimal uptake or loss of W during hydrothermal alteration of oceanic crust. Assuming a $10 \times$ increase in [W] for oceanic crust over MORB mantle (10% mantle melting), the slab crust enters the subduction zone with an elevated Xe/W ratio that is $10 \times$ to $10,000 \times$ greater than in MORB mantle. Tungsten and Xe are both mobile during subduction^{53–55}, but the relative efficiency of Xe and W removal from the slab during dehydration and melting is not known. Because of this uncertainty and the uncertainty on Xe/W ratios of plume mantle, we take the modern $^{182}\text{W}/^{184}\text{W}$ anomalies as the targets for modelling the isotopic evolution of plume mantle (no W dilution).

Calculation of W and Xe isotopic anomalies resulting from discrete stages of core formation. To identify scenarios where the short-lived isotopic signatures observed in plume and MORB mantle can be generated, we modelled the $^{182}\text{W}/^{184}\text{W}$ evolution and I depletion of mantle reservoirs that experienced single-stage core extraction at different P – T – X -timing conditions. We calculate core formation as a single-stage process to limit the number of free parameters within each calculation²³ and to facilitate P – T – X -timing comparisons between model outputs. Calculations of major-element chemistry resulting from core formation follow the approach of ref. 11, in which the following mass balance equation is solved:

$$[(\text{FeO})_x(\text{NiO})_y(\text{SiO}_2)_z(\text{Mg}_a\text{Al}_m\text{Ca}_n)\text{O}]_{\text{mantle}} + [\text{Fe}_a\text{Ni}_b\text{O}_c\text{Si}_d]_{\text{core}}$$

$$\rightarrow [(\text{FeO})_{x'}(\text{NiO})_{y'}(\text{SiO}_2)_{z'}(\text{Mg}_a'\text{Al}_m'\text{Ca}_n')\text{O}]_{\text{mantle}} + [\text{Fe}_{a'}\text{Ni}_{b'}\text{O}_{c'}\text{Si}_{d'}]_{\text{core}} \quad (14)$$

where a, b, c, u, m, n, x, y and z are defined by the composition of the starting material, and a', b', c', x', y', z' are unknowns determined on the basis of the partitioning of Ni, Si and O at the chosen P – T conditions. The starting material used for all calculations is an average of the ‘oxidized’ and ‘reduced’ impactor compositions given by ref. 11. A full description of the procedure used for the major-element calculations can be found in ref. 11. This approach has the advantage of coupling the P – T conditions of core formation with the mantle FeO content, an important constraint in our definition of successful model conditions. Partition

coefficients for I are calculated from equation (1). Note that the activity coefficient of Fe in liquid Fe alloy under the applicable T - X conditions is very close to 1 ($\ln(\gamma_{\text{met}}^{\text{Fe}}) = 0$) and can be neglected in applying equation (1) to natural conditions. Tungsten partitioning is given by ref. 18 and takes the following form:

$$\log(D_{\text{met/sil}}^{\text{W}}(\text{wt}\%)) = 1.85 - 6,728/T - 77 \times (P/T) + 3\log(D_{\text{met/sil}}^{\text{Fe}}(\text{wt}\%)) \quad (15)$$

A Monte Carlo approach is employed to search the parameter space, in which pressure and the timing of the core formation event are selected at random. Temperature at the chosen pressure is constrained to lie on the peridotite liquidus determined by ref. 22. The distribution of W and I between core and mantle is calculated from the given expressions for partitioning (equation (1) for I and using ref. 18 for W) applying the core-mantle chemistry calculated following ref. 11 (equation (14)). The evolution of the mantle W isotope composition subsequent to the core-forming event is then calculated using the following expression from ref. 62:

$$\mu_{\text{W(CHUR)}}^{182}(t) = Q_{\text{W}} \left(\frac{^{180}\text{Hf}}{^{182}\text{Hf}} \right)_{t_0} f^{\text{Hf/W}} (e^{-\lambda t_{2\text{stage}}} - e^{-\lambda t}) \quad (16)$$

where $\mu_{\text{W(CHUR)}}^{182}(t)$ is the parts per million deviation in the mantle W isotope composition at time t relative to CHUR, Q_{W} is related to $^{180}\text{Hf}/^{182}\text{Hf}$ of CHUR at $t = 0$, $f^{\text{Hf/W}}$ is the enrichment in Hf/W relative to CHUR that results from core formation, and $t_{2\text{stage}}$ corresponds to the time after Solar System formation of the core-forming event being considered.

Our initial goal was to use the model to determine the conditions of core-formation that are compatible with the composition of the MORB source reservoir, where this reservoir is approximately comprised of bulk silicate Earth. Models were considered successful when they matched $[\text{W}]^{63}$, $[\text{FeO}]$ and the isotopic composition of the bulk silicate Earth¹⁵ (Supplementary Table 4). Our Hf-W age determination of core segregation from MORB mantle occurs later in the accretion timeline than do previous determinations¹⁵. This difference stems from our requirement to satisfy $[\text{FeO}]_{\text{BSE}}$, in addition to $[\text{W}]_{\text{BSE}}$ and $^{182}\text{W}/^{184}\text{W}$ in the Hf-W calculation. The mutual satisfaction of $[\text{FeO}]_{\text{BSE}}$ and $[\text{W}]_{\text{BSE}}$ in a single-stage core formation framework is only possible at the lower end of $[\text{W}]_{\text{BSE}}$ estimates⁶³, which forces the Hf-W age to about 50 million years after the initiation of the Solar System (approximately the time of condensation of calcium-aluminium-rich inclusions, CAI) (Fig. 3). Satisfying only $[\text{W}]_{\text{BSE}}$ yields a Hf-W age of about 30 million years after CAI¹⁵, but the $[\text{FeO}]_{\text{BSE}}$ implied by this calculation exceeds observations. To calculate the value of $D_{\text{met/sil}}^{\text{I}}$ applicable to the core-forming event responsible for creating MORB source mantle, we employed the P - T conditions found by averaging the conditions of the successful models described above. Additional discussion of modelled $[\text{W}]_{\text{BSE}}$ is provided in the Methods subsection 'Collateral geochemical consequences'.

In the second step of the model we search the P -timing parameter space (T is fixed along the liquidus of ref. 22) to identify core extraction scenarios that produce a mantle reservoir with $\mu^{182}\text{W}$ values that are low or high relative to those of MORB source mantle by the amounts listed in Supplementary Table 4. In these model calculations, the measured $\mu^{182}\text{W}$ anomalies^{9,10} are the only constraint that must be satisfied in order to consider the result a success. These core-forming events are those considered further as candidates for the formation of plume mantle. Using the P - T conditions from these successful results, we calculate I partition coefficients and the corresponding depletions in mantle I that result from core extraction. Cases where the calculated I depletion is $\geq 2.8\times$ the value predicted for MORB source mantle are considered successful for both W and Xe isotopes and are plotted in Fig. 3a.

Dynamical simulations of planetary accretion suggest that the majority of volatile elements were delivered to Earth in the final 30% of growth⁶⁴. Relative I depletion between different sections of mantle (plume mantle and MORB mantle) in our model, however, is only a function of core formation. Our model explains I depletion in plume mantle without invoking early accretion of volatile-poor material. If the material that accreted to generate plume mantle was in fact volatile-depleted, it would reduce the pressure required for segregation of core metal from plume mantle (lower $D_{\text{met/sil}}^{\text{I}}$) and would allow for segregation later in Solar System history (Fig. 3a).

As noted above, the $^{129}\text{Xe}^*/^{136}\text{Xe}^*_{\text{Pu}}$ and $^{129}\text{Xe}/^{130}\text{Xe}_{\text{radiogenic}}$ offsets between plume and MORB mantle both require plume mantle I/Pu and I/Xe ratios to be approximately $3\times$ lower than MORB mantle upon formation. Given the differing volatilities of I, Pu and Xe, the I/Pu and I/Xe ratios of materials accreted to Earth should vary but to different degrees for materials with different volatile contents. This suggests that volatile delivery may have been roughly constant during the period overlapping plume mantle and MORB mantle core formation.

Plutonium, Hf and Xe are assumed not to enter the core in these calculations. Partitioning of Xe and other noble gases into the core has been studied by many groups and their uniform conclusion is that noble gases are not siderophile⁶⁵⁻⁶⁷, at least across the P - T - X conditions explored so far. Xenon partitioning data have only been reported between 0.5 GPa and 6 GPa and show a negative correlation with increasing pressure, with $D_{\text{met/sil}}^{\text{Xe}}$ values decreasing from about 0.1 to about 0.001 over this pressure range⁶⁶. We assume that Xe remains lithophile ($D_{\text{met/sil}}^{\text{Xe}} < 1$) across the P conditions considered here (up to 80 GPa). Partitioning of Pu during core formation has not been directly studied, but other actinides remain lithophile under the P - T - X conditions applicable to core formation in deep magma oceans⁶⁸.

We do not require that the calculated Xe closure age be the same as the timing of core extraction for MORB or plume mantle. Xenon closure ages depend on assumptions made for the abundance of I in bulk silicate Earth and assumptions regarding retention of Xe before complete closure. Moreover, both single-stage core formation and Xe closure calculations are idealized scenarios and do not yield ages with absolute chronological meaning. These assumptions make it difficult to directly compare the timing of core extraction and Xe closure ages.

We do not explore the independent effects on W partitioning related to oxygen fugacity (bulk oxygen content of the core-mantle system), S, C or temperatures removed from the mantle liquidus²². More reduced, C-rich, and S-poor core formation would make W more siderophile¹⁶⁻¹⁸ and vice versa. Combined, these effects would expand the P - T -timing space (Fig. 3a) for successful W isotopic solutions. **Collateral geochemical consequences.** High-pressure metal-silicate equilibrium would cause mantle to form with elevated abundances of moderately siderophile elements, FeO contents and oxygen fugacity^{11,37} (Fig. 2, Extended Data Fig. 3). Many plume-related materials contain these geochemical signals, including elevated Ni contents and elevated oxygen fugacity⁶⁸⁻⁷⁰ but (as emphasized in the main text) plume mantle Xe is dominated (about 90% of total Xe) by a recycled component²⁻⁴. Given that the relative efficiency of Xe recycling is probably lower than other less fluid-mobile elements it is plausible that geochemical signatures of accretion in plume mantle will be overprinted by recycled materials.

Tungsten is fluid-mobile under conditions relevant to subduction⁵⁵, and this behaviour may be crucial to preserving W isotopic anomalies²⁵. High P - T core formation generates mantle with elevated $[\text{W}]$ (Extended Data Fig. 3), and this elevated $[\text{W}]$ may also contribute to the long-term preservation of W anomalies. The fact that high P - T core formation generates mantle with elevated $[\text{W}]$ also leads to the prediction that the magnitude of the W anomalies will be correlated with $[\text{W}]$ in the mantle source (Extended Data Fig. 3).

According to the modelling presented here, the minimum core equilibration pressure that can generate the Xe and W isotopic signatures is about 55 GPa. Under these conditions, we predict mantle to form with $[\text{W}]$ about $5\times$ that of primitive MORB mantle (Extended Data Fig. 3). Partitioning behaviour for W follows from ref. 18. Higher-pressure core formation would produce mantle with higher $[\text{W}]$. At the highest pressures considered here (80 GPa), mantle will form with $[\text{W}]$ about $20\times$ compared to primitive MORB mantle. For Ni and Co, plume mantle will vary from about $2\times$ (55 GPa) to about $5\times$ primitive MORB values (80 GPa). Partitioning behaviour for Ni and Co follows from ref. 37. Primitive MORB source Ni and Co contents were calculated using the P - T - X identified as successful for matching bulk silicate Earth $[\text{W}]^{63}$ and $[\text{FeO}]$ (Supplementary Table 4), yielding bulk silicate Earth values for $[\text{Ni}]$ and $[\text{Co}]$ that are approximately $2\times$ lower than estimates⁷¹. Consequently, the relative trends, rather than the absolute values, of the MSE calculations are more useful in evaluating MSE behaviour and are presented here. We emphasize that these calculations are only applicable for plume mantle immediately following core formation. Components recycled into plume mantle will dilute any geochemical anomalies generated during the formation of plume mantle.

The effect of plume mantle on bulk silicate Earth $[\text{W}]$ is not clear, owing to uncertainty regarding the volume of plume mantle and $[\text{W}]$ in plume mantle. For example, plume mantle may be contained within modern large low-shear-velocity provinces, which comprise about 8% of the total mantle volume⁷². If this is true, modern plume mantle volume could range from around 8% of mantle volume to $< 1\%$. If plume mantle is 5% of mantle volume and contains around $20\times$ $[\text{W}]$ of the primitive MORB source (the high endmember case), plume mantle would contribute half of the bulk mantle budget. However, if plume mantle is 1% of the total mantle volume and contains $5\times$ $[\text{W}]$ of the primitive MORB source (the low endmember case), then plume mantle would contribute negligibly to the bulk mantle W budget. The most recent estimate⁶³ for BSE $[\text{W}]$ is 13 ± 10 p.p.b., so a doubling of bulk mantle $[\text{W}]$ is of similar order to current uncertainty estimates. Given that Ni and Co would be less enriched in plume mantle compared to W, the effect of plume mantle on Ni and Co bulk mantle budgets would be correspondingly smaller (Extended Data Fig. 3).

There are caveats to the expected correlation between the elevated $[\text{W}]$ and W anomalies. Given that Xe is 90% overprinted by recycled materials in plume

mantle^{2–4}, it is expected that the W anomalies have been diluted since their generation. Creating larger initial W anomalies in the mantle (as would be required if modern W anomalies have been diluted) requires earlier core formation or lower [W] in plume mantle. Lower [W] in the plume mantle could be achieved if plume mantle formed under more reduced conditions than considered here or under C-rich conditions^{16–18}. It is also possible that I was available at lower concentrations during plume mantle core formation. If true, this would allow plume mantle core formation to occur at pressures lower than 55 GPa, with correspondingly lower enrichments of W in plume mantle compared to MORB mantle. Thus, the approximately 5× increase of [W] in the plume mantle is only a minimum value for the specific conditions considered here. Nonetheless, elevated [W] associated with W anomalies is a prediction of the present model.

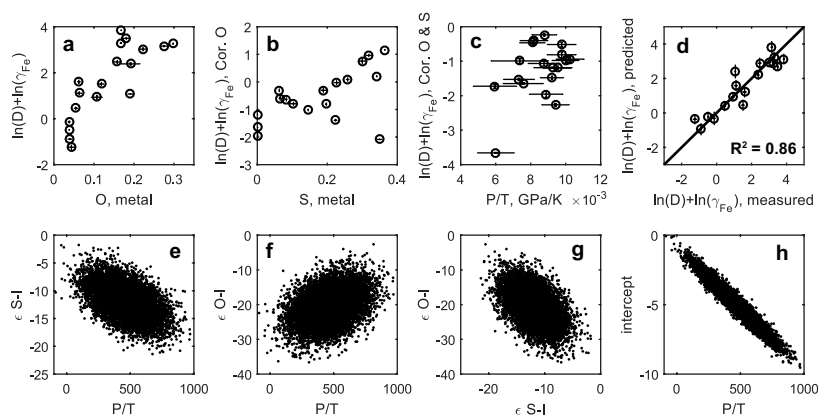
The pressure and timing of core formation for MORB mantle are constrained here by identifying the *P*–*t* timing space (*T* is fixed by the mantle liquidus of ref. 22) within a single-stage model that results in values for ¹⁸²W/¹⁸⁴W, [W], and [FeO] within the permissible range for the bulk silicate Earth (Supplementary Table 4). To calculate the *f*^{Hf/W} value associated with the evolution of ¹⁸²W/¹⁸⁴W in the MORB mantle following core formation (equation (16)), we assume⁶³ [Hf]_{BSE} = 280 p.p.b. Successful single-stage models are found for [W]_{BSE} ranging between 3.0 p.p.b. and 3.8 p.p.b., equating to a Hf/W_{BSE} range of 74–93. A lower Hf/W_{BSE} value of 18 is obtained by dividing best estimates of the Hf/U_{chondrite} against W/U_{BSE} (12.0 ± 4.2 and 0.65 ± 0.45, 2σ)^{63,73}. Calculating the corresponding uncertainty on the Hf/W_{BSE} ratio is complicated by the unknown covariance of Hf/U_{chondrite} and W/U_{BSE} uncertainties. Taking the extreme upper limit of Hf/U_{chondrite} and lower limit of W/U_{BSE} (negatively correlated uncertainties) yields an upper limit Hf/W_{BSE} ratio of 81. Thus, given the current parameterizations of W and FeO partitioning and ¹⁸²W/¹⁸⁴W constraints for CHUR and BSE^{11,15,18}, [W]_{BSE} and [FeO]_{BSE} can only be mutually satisfied in the single-stage core formation framework explored here at the upper limit of the potential Hf/W_{BSE} range.

To calculate the value of Hf/U_{chondrite} (12.0 ± 4.2) we average Hf/U measurements from ref. 73 for chondrites with ¹⁷⁶Hf/¹⁷⁷Hf ratios within error of the accepted value for chondrites of low metamorphic grade (0.282785 ± 0.000036, 2σ). We screen for accepted ¹⁷⁶Hf/¹⁷⁷Hf ratios in our averaging because elemental fractionations of Hf from U and Lu can occur during metamorphism on chondrite parent bodies⁷³, obscuring the bulk parent body U/Hf ratio. Fractionation of Hf from U and Lu is evidenced by a negative correlation between ¹⁷⁶Hf/¹⁷⁷Hf and Hf/U ratios within chondritic materials (not shown).

We emphasize that a single-stage core formation model is an endmember calculation. Expanding the present model to include multiple stages, non-liquidus temperatures, sulphur, carbon or changing oxygen content would enable greater overlap between the successful model Hf/W_{BSE} ratios and the geochemically constrained range of Hf/W_{BSE}. The offset between successful model Hf/W_{BSE} values and the best estimate of Hf/W_{BSE} does not change the essential result of the modelling, which is that mantle that experiences core formation, on average, under high pressures and early compared to the MORB mantle can evolve the range of W and Xe anomalies documented in plume rocks and will be predisposed to long-term preservation.

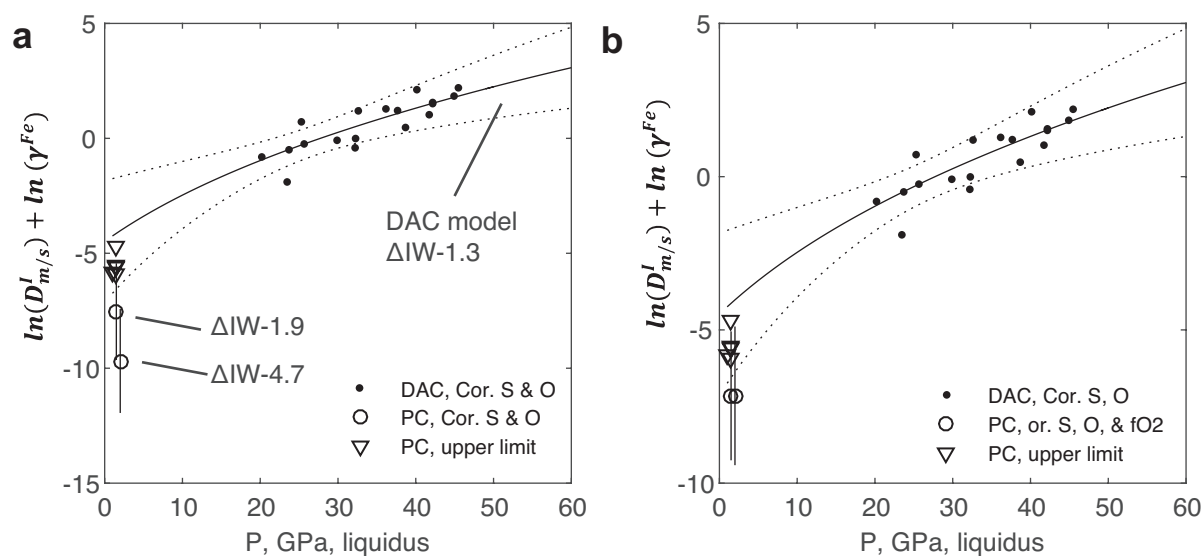
Data availability. The data supporting the findings of this study are available within the paper and its Supplementary Information.

31. Thibault, Y. & Walter, M. J. The influence of pressure and temperature on the metal-silicate partition coefficients of nickel and cobalt in a model C1 chondrite and implications for metal segregation in a deep magma ocean. *Geochim. Cosmochim. Acta* **59**, 991–1002 (1995).
32. Du, Z. *et al.* Using stepped anvils to make even insulation layers in laser-heated diamond-anvil cell samples. *Rev. Sci. Instrum.* **86**, 095103 (2015).
33. Akahama, Y. & Kawamura, H. Pressure calibration of diamond anvil Raman gauge to 310 GPa. *J. Appl. Phys.* **100**, 043516 (2006).
34. Prescher, C. & Prakapenka, V. B. DIOPTAS: a program for reduction of two-dimensional X-ray diffraction data and data exploration. *High Press. Res.* **35**, 223–230 (2015).
35. Andrault, D. *et al.* Thermal pressure in the laser-heated diamond-anvil cell: an X-ray diffraction study. *Eur. J. Mineral.* **10**, 931–940 (1998).
36. Tange, Y., Nishihara, Y. & Tsuchiya, T. Unified analyses for *P*–*V*–*T* equation of state of MgO: a solution for pressure-scale problems in high *P*–*T* experiments. *J. Geophys. Res. Solid Earth* **114** (B3), (2009).
37. Fischer, R. A. *et al.* High pressure metal-silicate partitioning of Ni, Co, V, Cr, Si, and O. *Geochim. Cosmochim. Acta* **167**, 177–194 (2015).
38. Du, Z. & Lee, K. K. High-pressure melting of MgO from (Mg, Fe) O solid solutions. *Geophys. Res. Lett.* **41**, 8061–8066 (2014).
39. Hirschmann, M. M., Baker, M. B. & Stolper, E. M. The effect of alkalis on the silica content of mantle-derived melts. *Geochim. Cosmochim. Acta* **62**, 883–902 (1998).
40. Liu, J., Li, J. & Ikuta, D. Elastic softening in Fe₇C₃ with implications for Earth's deep carbon reservoirs. *J. Geophys. Res. Solid Earth* **121**, 1514–1524 (2016).
41. Basu, S., Lahiri, A. K. & Seetharaman, S. Activity of iron oxide in steelmaking slag. *Metall. Mater. Trans. B* **39**, 447–456 (2008).
42. Japan Society for the Promotion of Science (JSPS) and the Nineteenth Committee on Steelmaking. *JSPS Steelmaking Data Sourcebook* 273–297 (Gordon and Breach Science Publishers, 1988).
43. Wade, J. & Wood, B. Core formation and the oxidation state of the Earth. *Earth Planet. Sci. Lett.* **236**, 78–95 (2005).
44. Ohnishi, S., Kuwayama, Y. & Inoue, T. Melting relations in the MgO–MgSiO₃ system up to 70 GPa. *Phys. Chem. Mineral.* **44**, 1–9 (2017).
45. Kilburn, M. & Wood, B. Metal-silicate partitioning and the incompatibility of S and Si during core formation. *Earth Planet. Sci. Lett.* **152**, 139–148 (1997).
46. Rubie, D. C., Gessmann, C. K. & Frost, D. J. Partitioning of oxygen during core formation on the Earth and Mars. *Nature* **429**, 58–61 (2004).
47. Dalou, C., Hirschmann, M. M., von der Handt, A., Mosenfelder, J. & Armstrong, L. S. Nitrogen and carbon fractionation during core-mantle differentiation at shallow depth. *Earth Planet. Sci. Lett.* **458**, 141–151 (2017).
48. Li, Y., Dasgupta, R., Tsuno, K., Monteleone, B. & Shimizu, N. Carbon and sulfur budget of the silicate Earth explained by accretion of differentiated planetary embryos. *Nat. Geosci.* **9**, 781–785 (2016).
49. Pepin, R. O. & Porcelli, D. Xenon isotope systematics, giant impacts, and mantle degassing on the early Earth. *Earth Planet. Sci. Lett.* **250**, 470–485 (2006).
50. Kendrick, M. *et al.* Seawater cycled throughout Earth's mantle in partially serpentinized lithosphere. *Nat. Geosci.* **10**, 222–228 (2017).
51. Touboul, M., Puchtel, I. S. & Walker, R. J. Tungsten isotopic evidence for disproportional late accretion to the Earth and Moon. *Nature* **520**, 530–533 (2015).
52. Kruijer, T. S., Kleine, T., Fischer-Gödde, M. & Sprung, P. Lunar tungsten isotopic evidence for the late veneer. *Nature* **520**, 534–537 (2015).
53. Heber, V. S., Brooker, R. A., Kelley, S. P. & Wood, B. J. Crystal-melt partitioning of noble gases (helium, neon, argon, krypton, and xenon) for olivine and clinopyroxene. *Geochim. Cosmochim. Acta* **71**, 1041–1061 (2007).
54. Smye, A. J. *et al.* Noble gases recycled into the mantle through cold subduction zones. *Earth Planet. Sci. Lett.* **471**, 65–73 (2017).
55. Bali, E., Keppler, H. & Audetat, A. The mobility of W and Mo in subduction zone fluids and the Mo–W–Th–U systematics of island arc magmas. *Earth Planet. Sci. Lett.* **351/352**, 195–207 (2012).
56. Matsuda, J. & Nagao, K. Noble gas abundances in a deep sea sediment core from eastern equatorial Pacific. *Geochem. J.* **20**, 71–80 (1986).
57. Staudacher, T. & Allegre, C. J. Recycling of oceanic crust and sediments—the noble gas subduction barrier. *Earth Planet. Sci. Lett.* **89**, 173–183 (1988).
58. Chavrit, D. *et al.* The contribution of hydrothermally altered ocean crust to the mantle halogen and noble gas cycles. *Geochim. Cosmochim. Acta* **183**, 106–124 (2016).
59. Kendrick, M. A., Honda, M. & Vanko, D. A. Halogens and noble gases in Mathematician Ridge meta-gabbros, NE Pacific: implications for oceanic hydrothermal root zones and global volatile cycles. *Contrib. Mineral. Petrol.* **170**, 43 (2015).
60. Kendrick, M. A., Scambelluri, M., Honda, M. & Phillips, D. High abundances of noble gas and chlorine delivered to the mantle by serpentinite subduction. *Nat. Geosci.* **4**, 807–812 (2011).
61. König, S., Münker, C., Schuth, S. & Garbe-Schönberg, D. Mobility of tungsten in subduction zones. *Earth Planet. Sci. Lett.* **274**, 82–92 (2008).
62. Harper, C. L. & Jacobsen, S. B. Evidence for ¹⁸²Hf in the early Solar System and constraints on the timescale of terrestrial accretion and core formation. *Geochim. Cosmochim. Acta* **60**, 1131–1153 (1996).
63. Arevalo, R. & McDonough, W. F. Tungsten geochemistry and implications for understanding the Earth's interior. *Earth Planet. Sci. Lett.* **272**, 656–665 (2008).
64. O'Brien, D. P., Walsh, K. J., Morbidelli, A., Raymond, S. N. & Mandell, A. M. Water delivery and giant impacts in the 'Grand Tack' scenario. *Icarus* **239**, 74–84 (2014).
65. Matsuda, J. *et al.* Noble gas partitioning between metal and silicate under high pressures. *Science* **259**, 788–790 (1993).
66. Sudo, M., Ohtaka, O. & Matsuda, J. Noble gas partitioning between metal and silicate under high pressures: the case of iron and peridotite. In *Noble Gas Geochemistry and Cosmochemistry* (ed. Matsuda, J.) 217–227 (Terra Scientifica, 1994).
67. Bouhifd, M., Jephcoat, A. P., Heber, V. S. & Kelley, S. P. Helium in Earth's early core. *Nat. Geosci.* **6**, 982–986 (2013).
68. Chidester, B. A., Rahman, Z., Righter, K. & Campbell, A. J. Metal-silicate partitioning of U: implications for the heat budget of the core and evidence for reduced U in the mantle. *Geochim. Cosmochim. Acta* **199**, 1–12 (2017).
69. Herzberg, C. *et al.* Nickel and helium evidence for melt above the core-mantle boundary. *Nature* **493**, 393–397 (2013).
70. Moussallam, Y. *et al.* The impact of degassing on the oxidation state of basaltic magmas: a case study of Kilauea volcano. *Earth Planet. Sci. Lett.* **450**, 317–325 (2016).
71. McDonough, W. F. & Sun, S. S. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
72. Cottaar, S. & Lekic, V. Morphology of seismically slow lower-mantle structures. *Geophys. J. Int.* **207**, 1122–1136 (2016).
73. Dauphas, N. & Pourmand, A. Hf–W–Th evidence for rapid growth of Mars and its status as a planetary embryo. *Nature* **473**, 489–492 (2011).



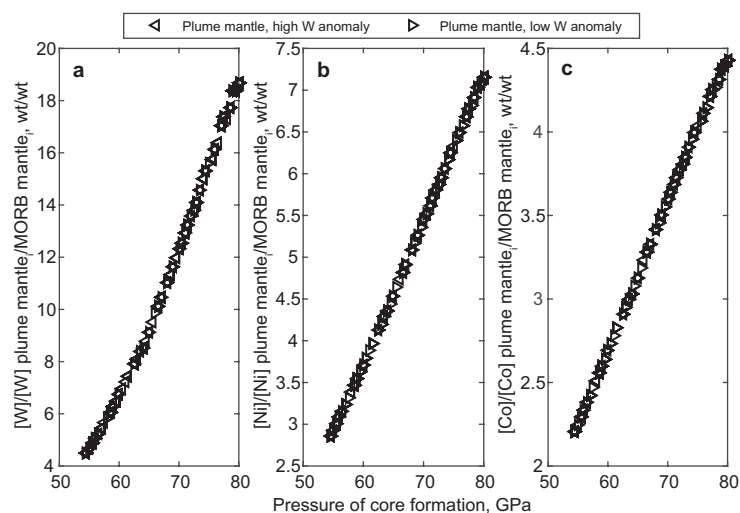
Extended Data Figure 1 | Parameterization of I partitioning between liquid Fe alloy and silicate liquid. **a**, Partitioning of I plotted against the O content (atomic) of metal phases. Oxygen content of the metal is the first parameter identified in the stepwise fitting approach. Higher O contents of metal are associated with greater partitioning of I into metal over silicate. **b**, Partitioning of I corrected to O-free metal plotted against S content of metal. The S content of the metal is the second parameter

identified in the stepwise fitting approach. **c**, Partitioning of I corrected to O- and S-free metal plotted against P/T . The P/T term is the third parameter identified in the stepwise fitting approach. **d**, A comparison of observed and predicted I partitioning. I partitioning is predicted using equation (1) ($R^2 = 0.86$). **e–h**, Covariance plots of parameterization. ‘Intercept’ refers to the constant term in equation (1).



Extended Data Figure 2 | Piston cylinder data comparison with DAC regression. **a**, A comparison of I partition coefficients determined in piston cylinder (PC) experiments and the I partition coefficient predicted from equation (1). Piston cylinder data are corrected to remove the effect of S and to the O contents of metal predicted along the mantle liquidus of ref. 22, allowing a direct comparison to the DAC data model. Upper-limit

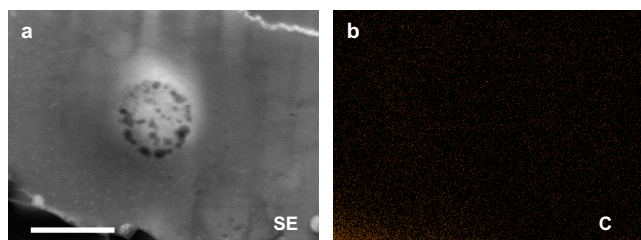
partition coefficients are consistent with the lower end of the uncertainty envelope. The difference in measured I partitioning coefficients in the piston cylinder series suggests that I partitioning is redox sensitive. **b**, Measured partition coefficients in piston cylinder series are corrected to $\Delta IW-1.3$ and offset in pressure for clarity. Error bars and the uncertainty envelope are 2σ .



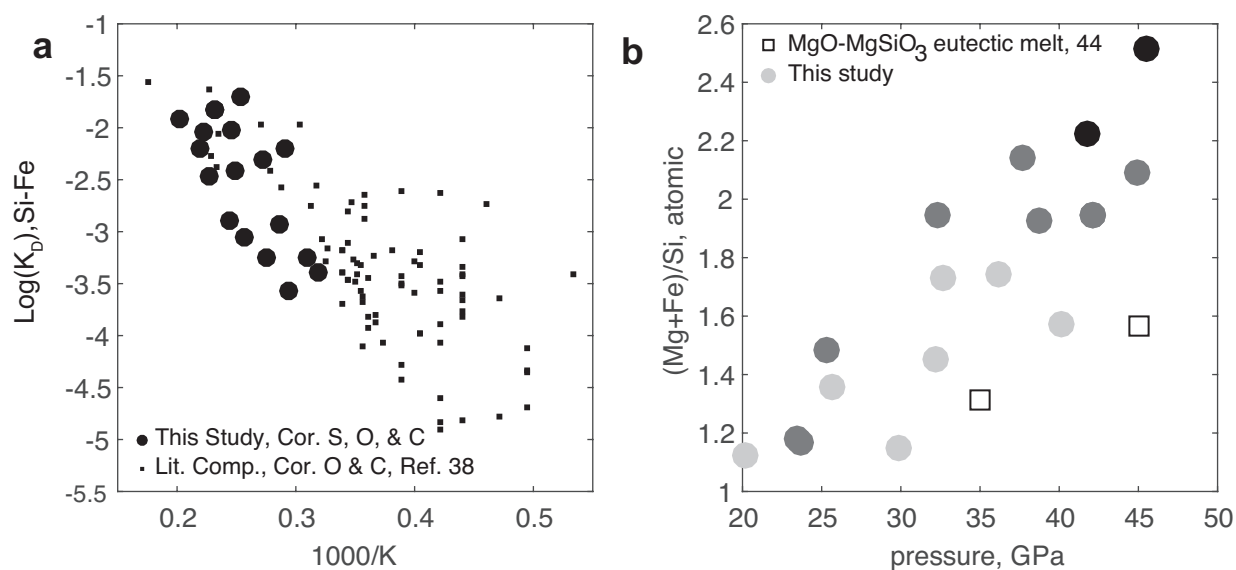
Extended Data Figure 3 | Calculated concentration of MSEs in plume mantle as a function of the pressure of core–mantle equilibrium.

W (a), Ni (b), and Co (c) concentrations in plume mantle all increase with increasing core–mantle equilibrium pressure. Results are normalized to the calculated concentration of MORB MSEs. MORB MSE concentrations are calculated using the average P – T – X conditions that satisfy BSE [W] and [FeO] (Supplementary Table 4). Temperatures are assumed to follow

the mantle liquidus of ref. 22. FeO abundances are calculated following ref. 11. W partitioning is from ref. 18. Ni and Co partitioning are from ref. 37. Plume mantle data points are plotted only for P – T – X conditions that satisfy W and Xe isotopic constraints (see Methods subsection ‘Calculation of W and Xe isotopic anomalies resulting from discrete stages of core formation’).

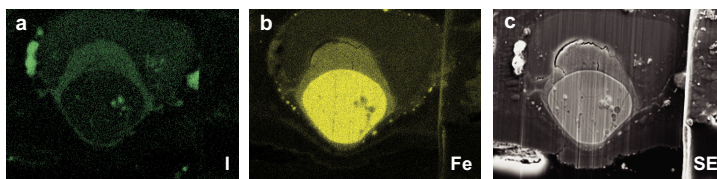


Extended Data Figure 4 | Imaging of silicate following analysis with low electron microprobe total. **a**, Secondary electron (SE) image of the DAC_I_EXP5 spot 4 silicate shows evidence of vesiculation in the area that was analysed. **b**, C map showing no obvious concentration of C local to vesiculation. The dark material in the lower left corner of **a** is diamond. Scale bar is 1 μm .



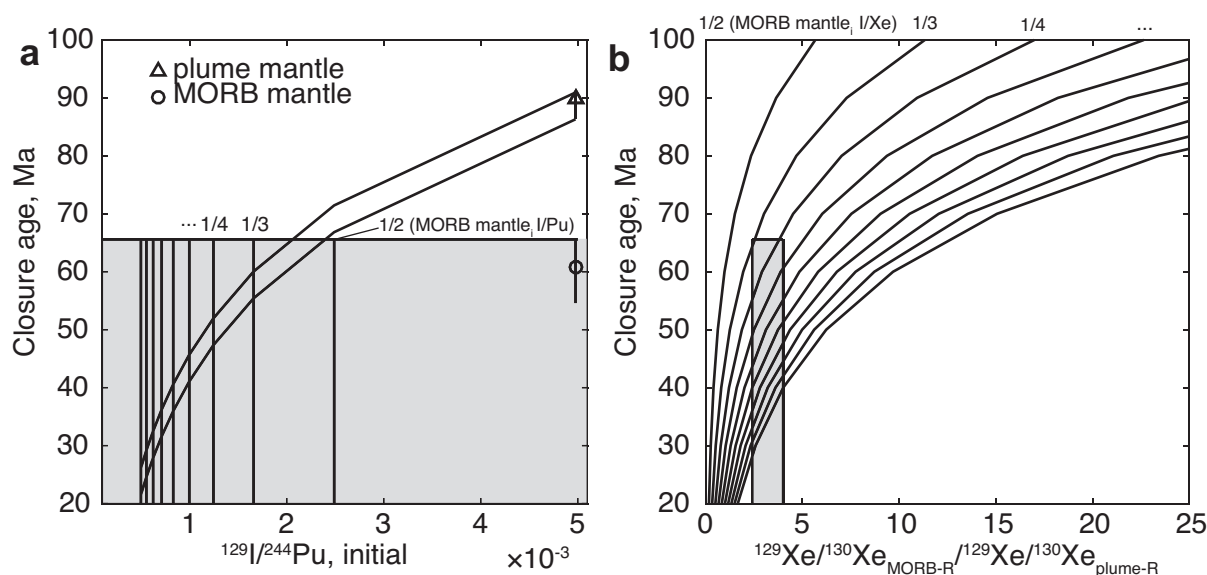
Extended Data Figure 5 | Demonstrations of DAC experiment equilibrium. **a**, Si-Fe exchange coefficient plotted against inverse temperature. Data from this study (circles) are corrected for S-O-C interactions with silicon. Squares are from the compilation of ref. 37 corrected from O-C interactions with silicon. **b**, $(\text{Mg}+\text{Fe})/\text{Si}$ ratios of

the silicate phase from this study (circles) and ref. 44 (squares) plotted against pressure. Symbols are grouped for Mg#, with darker symbols corresponding to lower Mg#. The offset of lower Mg# silicates within a pressure range to higher $(\text{Mg}+\text{Fe})/\text{Si}$ ratios and the similar pressure slopes within Mg# groupings support the accuracy of our pressure calculations.



Extended Data Figure 6 | Secondary electron image and EDS maps of experiment DAC_I_EXP9 spot 4 showing I mobility. The sequence of images shows the nature and distribution of I-rich materials that are mobilized to the surface of heating spots after storage in desiccators. **a**, Map of I distribution. I-rich materials are concentrated along the left and right edges of the heating spot, within isolated regions of the metal

and silicate phases, and surrounding the Fe alloy phase. **b**, Map of Fe distribution. The metallic phase is mantled by Fe-rich and I-rich material. This material is also rich in C and O (not shown). The I-rich material that is present along the edges of the heating spot is also enriched in Fe, C and O. **c**, Secondary electron image. I-rich materials are positive relief features. Images are 20 μm wide.



Extended Data Figure 7 | Parent-daughter ratio variations required to account for Xe isotopic variability between plume and MORB mantle.

a, Closure ages plotted against initial $^{129}\text{I}/^{244}\text{Pu}$ ratios. Closure ages are calculated using $^{129}\text{Xe}^*/^{136}\text{Xe}_{\text{Pu}}$ ratios from ref. 8 and equation (12). The MORB mantle $^{129}\text{I}/^{244}\text{Pu}$ ratio is fixed and derived from estimates for the bulk silicate Earth from refs 49 and 50. The grey shading delineates I-Pu-Xe closure ages that are equal to or less than that determined for MORB mantle. Vertical lines in the grey shaded area are fractional depletions of the I/Pu in plume mantle relative to MORB mantle, ordered sequentially. The ratios denote the fractional depletion of the I/Pu ratio for plume mantle relative to initial MORB mantle. An approximately $3 \times$ lower I/Pu

ratio for plume mantle results in equal closure ages for MORB and plume mantle. **b**, Closure ages plotted against the ratio of radiogenic components for $^{129}\text{Xe}/^{130}\text{Xe}$ within MORB and plume mantle. The upper edge of the grey shaded area delineates the upper limit of MORB mantle I-Pu-Xe closure. Left and right edges delineate the uncertainty on the ratio of radiogenic components observed in MORB and plume mantle. Curves are calculations for the ratio of radiogenic components of $^{129}\text{Xe}/^{130}\text{Xe}$ (denoted 'R'; see equation (13)), assuming different fractional depletions of the I/Xe ratio in plume mantle relative to initial MORB mantle (see fractions plotted along edge of graph). These calculations assume a MORB mantle closure age of 61 million years (Ma).

Monitoring T cell–dendritic cell interactions *in vivo* by intercellular enzymatic labelling

Giulia Pasqual¹, Aleksey Chudnovskiy¹, Jeroen M. J. Tas¹, Marianna Agudelo¹, Lawrence D. Schweitzer^{2,3}, Ang Cui^{2,3}, Nir Hacohen^{2,3} & Gabriel D. Victora¹

Interactions between different cell types are essential for multiple biological processes, including immunity, embryonic development and neuronal signalling. Although the dynamics of cell–cell interactions can be monitored *in vivo* by intravital microscopy¹, this approach does not provide any information on the receptors and ligands involved or enable the isolation of interacting cells for downstream analysis. Here we describe a complementary approach that uses bacterial sortase A-mediated cell labelling across synapses of immune cells to identify receptor–ligand interactions between cells in living mice, by generating a signal that can subsequently be detected *ex vivo* by flow cytometry. We call this approach for the labelling of ‘kiss-and-run’ interactions between immune cells ‘Labelling Immune Partnerships by SorTagging Intercellular Contacts’ (LIPSTIC). Using LIPSTIC, we show that interactions between dendritic cells and CD4⁺ T cells during T-cell priming *in vivo* occur in two distinct modalities: an early, cognate stage, during which CD40–CD40L interactions occur specifically between T cells and antigen-loaded dendritic cells; and a later, non-cognate stage during which these interactions no longer require prior engagement of the T-cell receptor. Therefore, LIPSTIC enables the direct measurement of dynamic cell–cell interactions both *in vitro* and *in vivo*. Given its flexibility for use with different receptor–ligand pairs and a range of detectable labels, we expect that this approach will be of use to any field of biology requiring quantification of intercellular communication.

LIPSTIC is based on proximity-dependent labelling across cell–cell interfaces using the *Staphylococcus aureus* transpeptidase sortase A (SrtA). SrtA covalently transfers a substrate containing the sorting motif ‘LPXTG’ to a nearby N-terminal oligoglycine² (Extended Data Fig. 1). For LIPSTIC, a ligand and receptor of interest are genetically fused to either SrtA or a tag that consists of five N-terminal glycine residues (G5). Addition of a SrtA substrate (for example, an LPETG peptide linked at the N terminus to a detectable label, such as biotin or a fluorophore) leads to loading of this peptide onto SrtA on the donor cell via the formation of an acyl intermediate. When a ligand and receptor interact, SrtA catalyses the transfer of the substrate onto the G5-tagged receptor. After cells separate, the interaction history is revealed by the presence of the label on the surface of the G5-expressing cell (Fig. 1a). To ensure that labelling occurs specifically as a readout of the ligand–receptor interaction—rather than being driven by the affinity of SrtA to G5—we used an engineered SrtA variant with a 13-fold lower affinity for G5 compared to wild-type SrtA (K_m of engineered SrtA = $1,830 \pm 330 \mu\text{M}$ compared to a $K_m = 140 \pm 30 \mu\text{M}$ for wild-type SrtA)³. This affinity is orders of magnitude lower than most receptor–ligand interactions involved in immune function^{4–7}.

To test this system, we transfected two populations of HEK293T cells separately with either G5–CD40 or CD40L–SrtA, mixed the two populations in the presence of the biotinylated SrtA substrate (biotin–LPETG) for 30 min, and then analysed the cells by flow

cytometry and western blot. To determine specificity, G5–CD40 cells were also incubated with HEK293T cells transfected with SrtA fused to a CD40L variant carrying two point mutations that strongly impair binding to CD40^{8,9} (CD40L*–SrtA, Extended Data Fig. 2) or with untargeted SrtA anchored to the cell surface by the transmembrane domain of PDGFR (SrtA–PDGFR) (Fig. 1b, c). Flow cytometric analysis showed that G5–CD40⁺ cells were biotinylated efficiently only when incubated with cells expressing wild-type CD40L–SrtA (Fig. 1d). Western blotting confirmed that LIPSTIC labelling occurred via covalent modification of G5–CD40 (Fig. 1e). Specific intercellular labelling was also achieved with other ligand–receptor pairs that are involved in immune cell interactions and neuronal signalling, indicating that LIPSTIC can be used to analyse a variety of molecular interactions (Fig. 1f–h). To visualize the dynamics of LIPSTIC labelling *in vitro*, we imaged interactions between B cells transduced with G5–CD40 and CD4⁺ T cells transduced with CD40L–SrtA and preloaded with Alexa Fluor 647–LPETG. Substrate transfer between T and B cells was observed within minutes of interaction and at the interacting surface (Extended Data Fig. 3 and Supplementary Video 1). We conclude that LIPSTIC is an efficient, specific and versatile method that is able to label receptor–ligand interactions across cells *in vitro*, suitable for use with multiple receptor–ligand pairs and for detection by both flow cytometry and microscopy.

To determine whether LIPSTIC can function *in vivo* and at endogenous levels of receptor and ligand expression, we generated mice carrying *Cd40*^{G5} and *Cd40lg*^{SrtA} alleles targeted to their endogenous loci (Extended Data Fig. 4). Expression of G5–CD40 was made constitutive, whereas expression of CD40L–SrtA was designed to occur only after Cre-mediated excision of a translational stop cassette, in order to specify the SrtA⁺ donor cell population. To measure LIPSTIC labelling during antigen-specific interactions between T cells and antigen-presenting cells, we crossed *Cd40lg*^{SrtA} to CD4-Cre and to OT-II TCR mice, which express a T-cell receptor specific for the chicken ovalbumin (OVA) peptide OVA_{323–339} (we refer to this strain as OT-II–SrtA). We co-cultured OT-II–SrtA CD4⁺ T cells for 6 h with *Cd40*^{G5/G5} splenic dendritic cells treated either with OVA_{323–339} or with a control LCMV-GP_{61–80} peptide and added the biotinylated substrate during the final 20 min of culture (Fig. 2a). Efficient intercellular labelling only occurred when dendritic cells were treated with the cognate peptide, which correlated with induction of CD40L–SrtA expression on T cells. Dendritic cell labelling was strongly inhibited by addition of a CD40L-blocking antibody, confirming that the CD40–CD40L interaction is required (Fig. 2b, c). LIPSTIC labelling was dose-responsive over a six-log range of OVA peptide concentrations (Fig. 2d, e). Co-culture of OT-II–SrtA CD4⁺ T cells with two *Cd40*^{G5/G5} dendritic cell populations separately pulsed with either the OVA_{323–339} or control LCMV-GP_{61–80} peptide showed that labelling was restricted to dendritic cells loaded with the cognate antigen, also across a wide range of antigen doses (Fig. 2f–h). Whereas labelling of cognate dendritic cells increased

¹Laboratory of Lymphocyte Dynamics, The Rockefeller University, 1230 York Avenue, New York, New York, USA. ²Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. ³Center for Cancer Research, Massachusetts General Hospital, Department of Medicine, Boston, Massachusetts, USA.

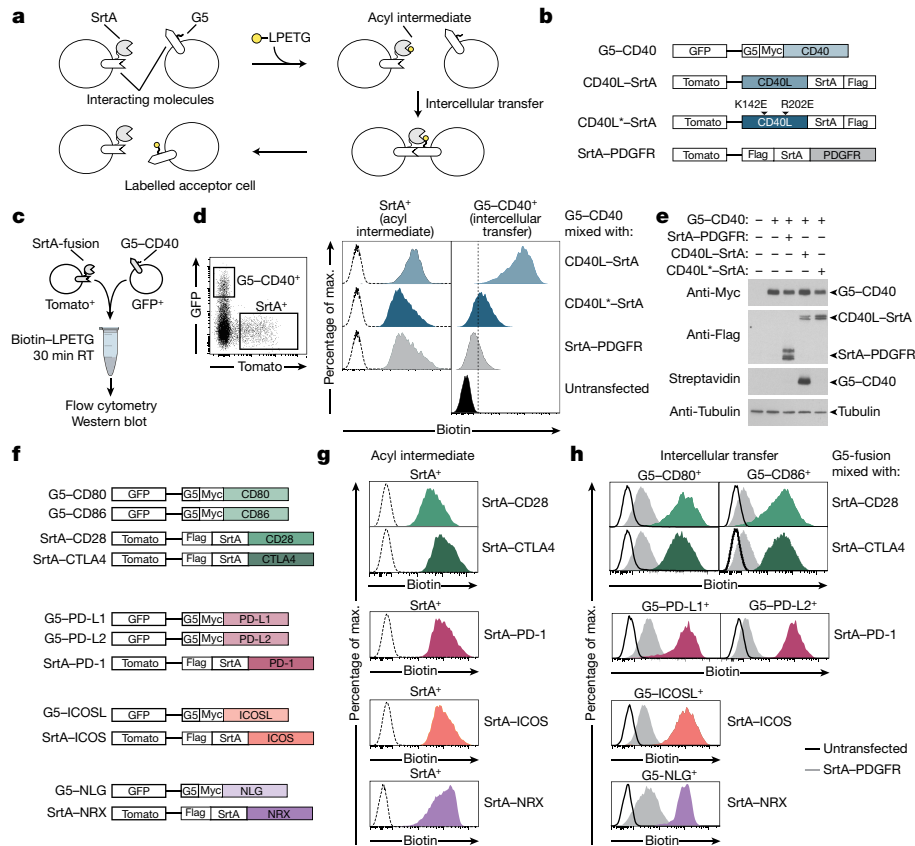


Figure 1 | Using LIPSTIC to track ligand–receptor interactions.

a, Schematic representation of the LIPSTIC approach. **b**, Constructs used in **c–e**. All constructs express a bicistronic gene that encodes a fluorescent reporter protein. **c**, Experimental setup to analyse intercellular labelling in transfected HEK293T cells. RT, room temperature. **d**, Gating strategy and histograms showing biotin staining in SrtA⁺ cells (left column, indicating the formation of the acyl intermediate) and G5–CD40⁺ cells (right column, indicating intercellular transfer). Histograms with dashed lines represent untransfected cells. **e**, Western blot showing expression of

G5–CD40 (anti-Myc), SrtA fusion constructs (anti-Flag) and intercellular labelling (Streptavidin). Tubulin is used as a loading control. **f**, Constructs used in **g, h**. NLG, neuroligin; NRX, neurexin. **g, h**, Biotin staining in SrtA⁺ cells (acyl intermediate) and G5⁺ cells (intercellular transfer). Histograms with dashed lines represent untransfected cells. Histograms with solid lines and grey histograms represent G5⁺ cells mixed with untransfected and SrtA–PDGFR donor cells, respectively. Data are representative of three independent experiments.

when the substrate was incubated for a longer time, labelling of control dendritic cells was negligible even when the substrate was present for the full 6 h of co-culture (Extended Data Fig. 5a–c). LIPSTIC was also capable of specifically identifying B cells that were productively engaged with antigen, as determined by co-culture of antigen-specific and polyclonal B cells with OT-II–SrtA CD4⁺ T cells (Extended Data Fig. 5d–g). Therefore, LIPSTIC labelling in short-term *ex vivo* priming experiments is dependent on interactions between receptor and ligand, dose-responsive across a wide range of antigen concentrations, and specific to target cells displaying cognate antigens. Of note, although SrtA–CD40L was capable of stimulating B-cell activation when expressed on HEK293T cells (Extended Data Fig. 2c), B-cell activation by CD40L–SrtA CD4⁺ T cells was impaired both *ex vivo* and *in vivo* when compared to activation by T cells expressing wild-type CD40L, indicating that signalling by CD40L is partly compromised (Extended Data Fig. 6a, b). This impairment was also seen in mice not carrying the CD4–Cre transgene, which expressed a construct that only had a translated LoxP site added to the C terminus of CD40L (Extended Data Fig. 6b); this impairment therefore more likely represents a specific feature of the CD40L molecule than a general property of SrtA–fusion proteins. Nevertheless, experiments using dendritic cells as antigen-presenting cells showed no measurable effect of the *Cd40lg*^{SrtA} allele on T-cell proliferation, indicating that the overall kinetics of T-cell priming are not affected by CD40L insufficiency (Extended Data Fig. 6c); we therefore used interactions between T cells and dendritic cells to characterize LIPSTIC labelling *in vivo*.

To determine whether LIPSTIC can be used *in vivo*, we used a well-established T-cell priming model in which OVA_{323–339}-treated *Cd40*^{G5/G5} dendritic cells were injected subcutaneously into the footpad of recipient mice, followed 18 h later by intravenous transfer of OT-II–SrtA CD4⁺ T cells¹⁰. We delivered the LIPSTIC substrate to the popliteal lymph node (PLN) by footpad injection of a total of 300 nmol of biotin–LPETG over six injections between 10 and 12 h after T-cell transfer (Fig. 3a); T cells are engaged in long-lived interactions with antigen-bearing dendritic cells at this time, as determined by intravital imaging¹⁰. Flow cytometry of PLN cells showed efficient LIPSTIC labelling of transferred dendritic cells, which was dependent on T-cell expression of CD40L–SrtA and sensitive to treatment with a CD40L-blocking antibody (Fig. 3b, c and Extended Data Fig. 7a). Background labelling was negligible in all assayed cell populations (Extended Data Fig. 7b). To further confirm the dependence of LIPSTIC labelling on CD40–CD40L interaction, we took advantage of the observation that, in the absence of a *Cd40*^{G5} allele, endogenous N-terminal glycines on the cell surface can function as low-efficiency acceptors for the SrtA substrate¹¹ (Extended Data Fig. 7c, d). Such labelling was completely absent when Ag-loaded dendritic cells were deficient in *Cd40*, again showing that CD40–CD40L engagement is essential for labelling (Extended Data Fig. 7e, f). Analysis of the kinetics of substrate clearance from labelled cells showed that a fraction of the label was still detectable at 4 and 8 h after substrate injection (Extended Data Fig. 7g–k).

To measure the interaction between CD4⁺ T cells and endogenous dendritic cells after immunization, we adoptively transferred

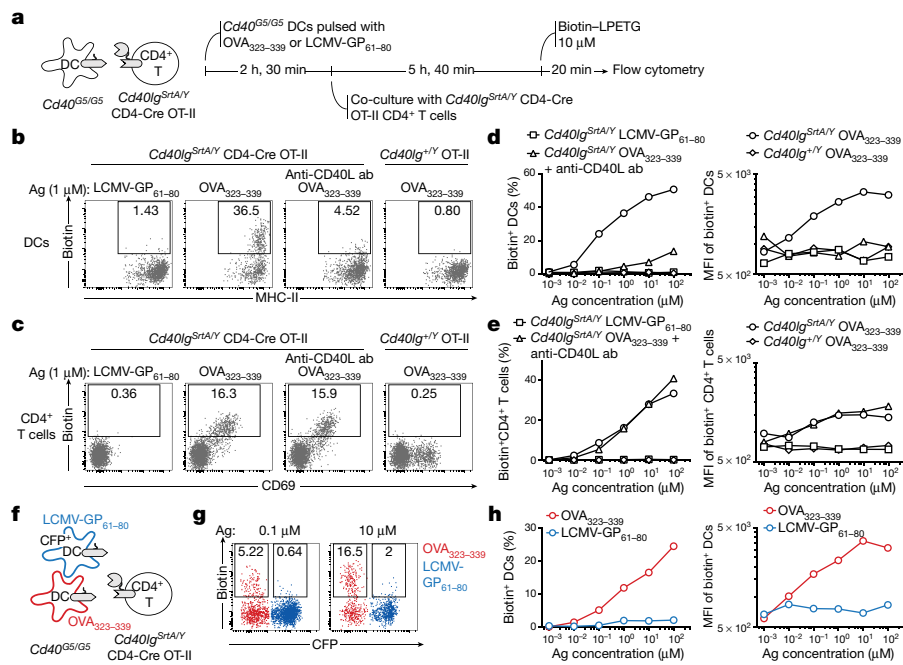


Figure 2 | LIPSTIC labelling of CD40-CD40L interactions *ex vivo*. **a**, Experimental setup for **b-e**. *Cd40lg^{SrtA/Y}*, male mice with only one copy of the *Cd450lg^{SrtA}* allele (*Cd40lg* is an X-chromosome-linked gene). **b**, Flow cytometry for biotin labelling of dendritic cells treated with 1 μM of the indicated peptide (intercellular transfer). **c**, Flow cytometry for biotin labelling of CD4⁺ T cells (acyl intermediate). **d**, **e**, Percentage (left) and median fluorescence intensity (MFI, right) of biotin⁺ dendritic cells (**d**) and CD4⁺ T cells (**e**). **f**, Experimental setup for **g**, **h**; timeline as in **a**. **g**, Flow cytometric analysis of dendritic cells treated with 0.1 or 10 μM of the indicated peptides showing biotin labelling (intercellular transfer). **h**, Percentage (left) and MFI (right) of biotin⁺ dendritic cells gated as in **g**. All data are representative of three independent experiments.

OT-II-SrtA CD4⁺ T cells into *Cd40^{G5/G5}* hosts and performed *in vivo* LIPSTIC labelling at different times after footpad injection of 10 μg of OVA in an alum adjuvant (Fig. 3d). LIPSTIC labelling was observed as early as 24 h after immunization on a small fraction of MHC-II^{hi} dendritic cells, which are likely to be pioneer antigen-presenting cells that drive the initiation of the T cell response

in the draining lymph node. The fraction of labelled dendritic cells increased over time, peaking at 10–15% of all dendritic cells at 72 h after immunization (Fig. 3e–f and Extended Data Fig. 7). Phenotypic analysis showed that labelling was restricted to MHC-II^{hi} dendritic cells, which were mostly CD11b⁺. Labelling of XCR1⁺ dendritic cells was a rare event, and was observed consistently—albeit

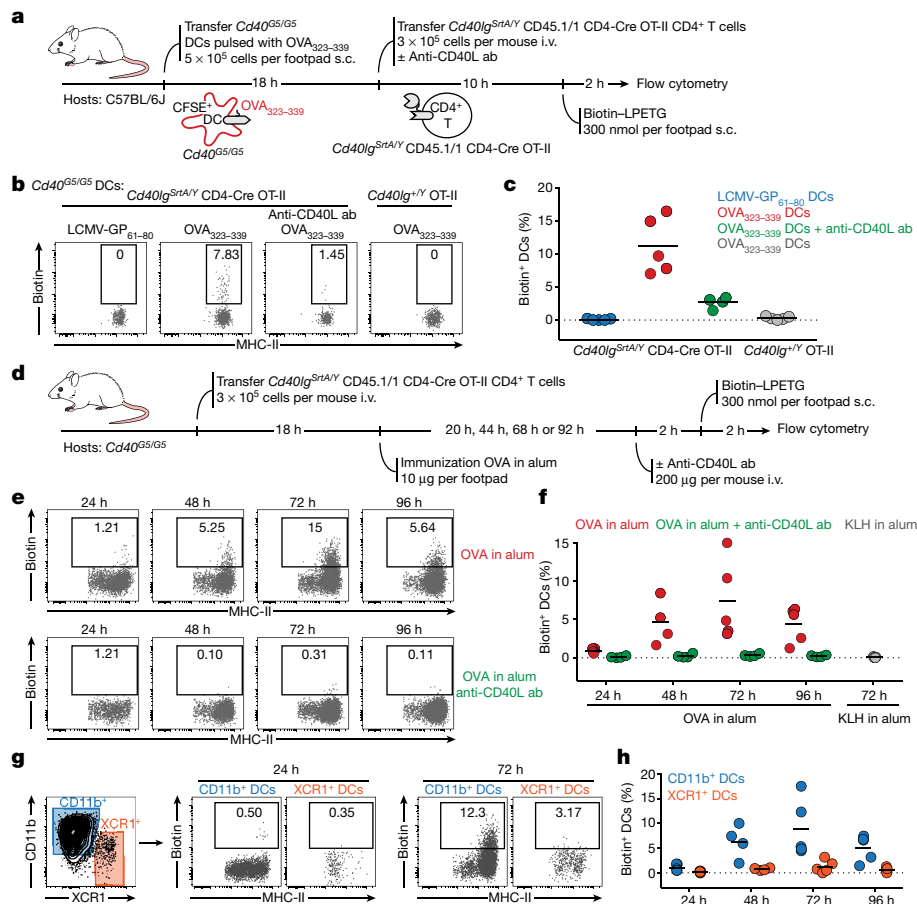


Figure 3 | LIPSTIC enables monitoring of CD40-CD40L interactions between T cells and dendritic cells *in vivo*. **a**, Experimental setup for **b**, **c**. CD45.1 encoded by *Ptprc*; homozygotes are indicated as CD45.1/1. **b**, Flow cytometric analysis of PLN cells showing biotin labelling of transferred *Cd40^{G5/G5}* dendritic cells. **c**, Percentage of biotin⁺ dendritic cells among transferred dendritic cell populations gated as in **b**. **d**, Experimental setup for **e-h**. **e**, Flow cytometric analysis of PLN cells showing biotin labelling of endogenous dendritic cells at different times after immunization in mice left untreated (top) or treated with a CD40L-blocking antibody (bottom). **f**, Percentage of biotin⁺ dendritic cells among endogenous dendritic cells gated as in **e**. **g**, Flow cytometric analysis of PLN cells showing biotin labelling among endogenous CD11b⁺ or XCR1⁺ dendritic cells at 24 or 72 h after immunization. **h**, Percentage of biotin⁺ dendritic cells among endogenous CD11b⁺ or XCR1⁺ dendritic cells gated as in **g**. **c**, **f**, **h**, Each symbol represents one mouse; bars indicate the mean. Data are pooled from two independent experiments. i.v., intravenous injection; s.c., subcutaneous injection.

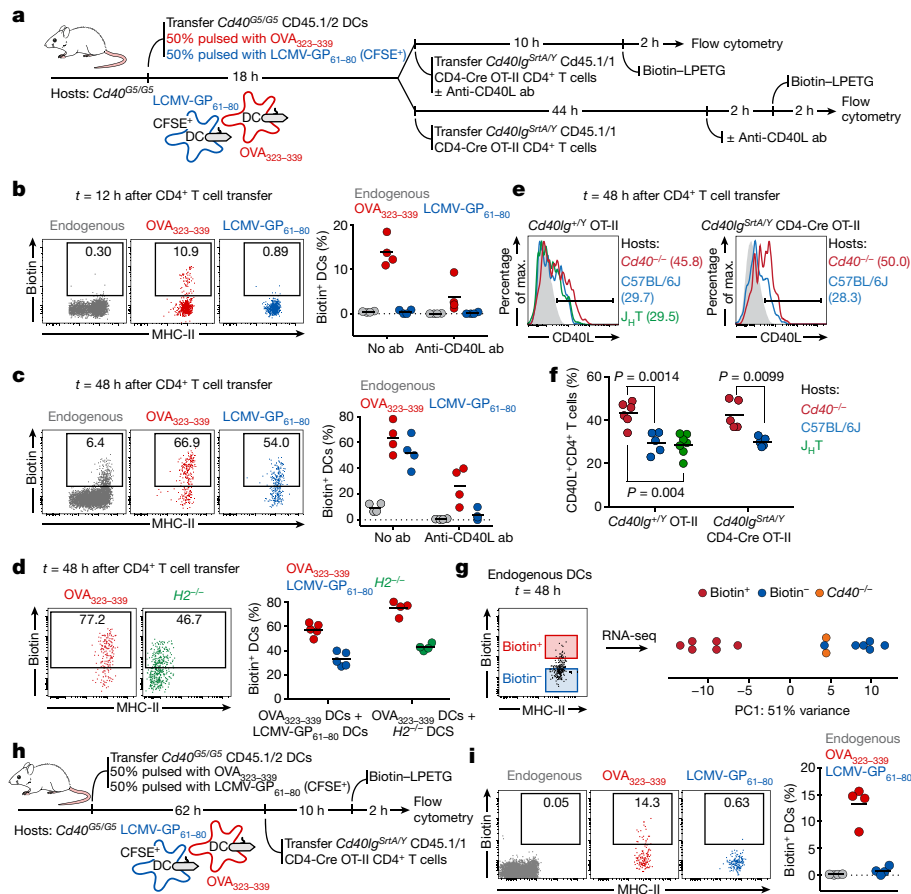


Figure 4 | Different modalities of the CD40-CD40L interaction between CD4⁺ T cells and dendritic cells *in vivo*. **a**, Experimental setup for **b**, **c**, **d**, **e**, Flow cytometric analysis of PLN cells showing biotin labelling of endogenous and transferred dendritic cells 12 h (**b**) or 48 h (**c**) after T-cell transfer. **d**, Flow cytometry of PLN cells showing biotin labelling of transferred dendritic cells 48 h after T-cell transfer. Experimental setup as in **a**, except that bystander dendritic cells are H2^{-/-} cells. **e**, CD40L expression in activated CD4⁺ T cells. Histograms show CD40L surface staining in *Cd40^{G5/G5}* OT-II (left) or *Cd40^{G5/G5}* CD4-Cre OT-II (right) CD69⁺ CD4⁺ T cells. Data are representative of two independent experiments. **f**, Percentage of CD69⁺ CD4⁺ T cells positive for CD40L. One-way ANOVA with Tukey's post hoc test and unpaired two-tailed Student's *t*-test were used for statistical analysis.

at low levels—only at 72 h after immunization, in line with previous reports that used intravital imaging and histocytometry¹² (Fig. 3g, h). We conclude that LIPSTIC can be used to follow the dynamics of CD40-CD40L contacts between T cells and dendritic cells *in vivo*, with a sufficient signal-to-noise ratio to detect rare and low-intensity interactions.

The finding that the CD40-CD40L interaction between T cells and dendritic cells peaks at 72 h after immunization (Fig. 3f), in addition to previous studies that have suggested that CD40L may, under certain circumstances, engage its receptor in the absence of antigen presentation¹³⁻¹⁵, led us to hypothesize that LIPSTIC labelling later in the response may reflect non-cognate interactions between T cells and dendritic cells taking place during the motile 'phase 3' of T-cell priming¹⁰. To test this hypothesis, we co-transferred into *Cd40^{G5/G5}* hosts two populations of dendritic cells treated independently with either OVA₃₂₃₋₃₃₉ (the cognate population) or LCMV-GP₆₁₋₈₀ (the bystander population), followed by OT-II-SrtA CD4⁺ T cells (Fig. 4a). Whereas LIPSTIC labelling at 12 h after T-cell transfer was detected only on cognate dendritic cells, specificity was lost at 48 h, when both transferred and endogenous bystander dendritic cells were robustly labelled (Fig. 4b, c and Extended Data Fig. 8a). To verify that bystander

labelling was truly non-cognate—as opposed to resulting from transfer of antigenic peptide between dendritic cell populations—we performed identical co-transfer experiments, but using H2^{-/-} dendritic cells as bystanders. LIPSTIC labelling of H2^{-/-} dendritic cells at late time points was indistinguishable from that of MHC-II-sufficient bystanders under the same conditions (Fig. 4d). Non-cognate LIPSTIC labelling of bystander dendritic cells at later time points was also observed after OVA immunization of haematopoietic chimaeras reconstituted with 80% *Cd40^{G5/G5}* and 20% *Cd40^{G5/G5}*; H2^{-/-} bone marrow (Extended Data Fig. 8c-e), and during *ex vivo* priming experiments analogous to those described in Fig. 2 (Extended Data Fig. 9). Thus, CD40L-CD40 LIPSTIC labelling during late stages of T-cell priming is not restricted to dendritic cells presenting the cognate antigen, in three distinct priming models.

To confirm non-cognate CD40-CD40L interactions in a system independent of LIPSTIC, we took advantage of the observation that CD40-dependent downregulation of surface CD40L on T cells can be used as a surrogate reporter for the CD40-CD40L interaction *in vivo*¹⁴. We injected wild-type OVA₃₂₃₋₃₃₉-treated dendritic cells into either wild-type or *Cd40^{-/-}* hosts, transferred OT-II CD4⁺ T cells one day later, then analysed CD40L expression at 48 h after

T-cell transfer. Whereas downregulation of CD40L could be observed in T cells transferred into wild-type hosts, this was not the case for T cells transferred into *Cd40*^{-/-} hosts, despite the presence of CD40 on the transferred Ag-loaded dendritic cells. CD40L downregulation was comparable in wild-type and in B-cell-deficient (J_HT) hosts, indicating that B cells do not contribute to CD40L downregulation (Fig. 4e, f). Thus, CD40L–CD40 interactions between T cells and with non-B cell antigen-presenting cells that are not loaded with antigen can downregulate surface CD40L on the T cell, confirming the interactions between activated T cells and bystander dendritic cells revealed by LIPTSIC. Moreover, similar downregulation of CD40L was observed in OT-II–SrtA T cells, indicating that the SrtA fusion does not prevent the downregulation of CD40L after it engages CD40 (Fig. 4e–f).

Gene expression profiling of bystander biotin⁺ and biotin⁻ host dendritic cells revealed clear differences between these two populations. Principal component analysis detected a major component (accounting for 51% of total variance) for which biotin⁺ dendritic cells were clearly separated from biotin⁻ dendritic cells, which in turn resembled bystander dendritic cells from mice lacking CD40 (Fig. 4g), a conclusion also supported by hierarchical clustering (Extended Data Fig. 10c). Differential expression analysis identified 788 genes that differed significantly between conditions (fold change > 2 and false-discovery rate < 0.05) (listed in Extended Data Fig. 10d and Supplementary Table1). We conclude that bystander interactions between T cells and dendritic cells are associated with marked changes in gene expression, and CD40 ligation potentially has a role in these alterations.

Finally, to determine whether the change in the pattern of interaction between T cells and dendritic cells over time is due to a change in the properties of T cells or of dendritic cells, we repeated the OVA_{323–339}/LCMV-GP_{61–80} dendritic cell co-transfer experiment but this time delaying T-cell transfer to 62 h after dendritic cell injection, so that, at the time of labelling, dendritic cells had been in the host for 70 h but T cells for only 12 h (Fig. 4h). This rescued the specificity of T-cell interactions, in that only dendritic cells treated with OVA_{323–339} peptide were labelled (Fig. 4i). Thus, newly primed T cells retain their specificity even when antigen-bearing dendritic cells have been present for several days. We conclude that CD40–CD40L interactions between CD4⁺ T cells and dendritic cells proceed in two stages. Initially, CD40L signals from arrested T cells are delivered specifically to antigen-loaded dendritic cells that are priming the response. This is followed by an antigen-independent stage in which motile, activated T cells are capable of interacting via CD40L even with dendritic cells that are not presenting the cognate antigen.

We introduce LIPSTIC, a novel system for labelling cell–cell interactions enzymatically both *in vitro* and *in vivo*. Although similar approaches have been proposed previously^{16–18}, our system has a number of unique features: first, the use of a mutated version of SrtA with low affinity for the N-terminal oligoglycine makes SrtA less likely to be the driver of the cell–cell interaction and more likely to operate as a readout of interactions driven by high-affinity receptor–ligand pairs. Second, SrtA uses a peptide substrate that is easily synthesized and can be linked to a wide variety of detectable labels, including genetically encoded fluorescent proteins or epitope tags². Third, and most importantly, SrtA substrates can be readily administered to live animals, allowing us to detect and isolate cells based on their history of intercellular interactions *in vivo*. Given the importance of such interactions to immunology and other fields of biomedical science, we expect this technology will be widely useful to biologists in general, representing a useful complement to intravital microscopy.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 June; accepted 8 December 2017.

Published online 17 January 2018.

- Cahalan, M. D. & Parker, I. Choreography of cell motility and interaction dynamics imaged by two-photon microscopy in lymphoid organs. *Annu. Rev. Immunol.* **26**, 585–626 (2008).
- Popp, M. W. & Ploegh, H. L. Making and breaking peptide bonds: protein engineering using sortase. *Angew. Chem. Int. Ed. Engl.* **50**, 5024–5032 (2011).
- Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl Acad. Sci. USA* **108**, 11399–11404 (2011).
- Haswell, L. E., Glennie, M. J. & Al-Shamkhani, A. Analysis of the oligomeric requirement for signaling by CD40 using soluble multimeric forms of its ligand, CD154. *Eur. J. Immunol.* **31**, 3094–3100 (2001).
- Ghiotto, M. et al. PD-L1 and PD-L2 differ in their molecular mechanisms of interaction with PD-1. *Int. Immunol.* **22**, 651–660 (2010).
- van der Merwe, P. A., Bodian, D. L., Daenke, S., Linsley, P. & Davis, S. J. CD80 (B7-1) binds both CD28 and CTLA-4 with a low affinity and very fast kinetics. *J. Exp. Med.* **185**, 393–404 (1997).
- Greene, J. L. et al. Covalent dimerization of CD28/CTLA-4 and oligomerization of CD80/CD86 regulate T cell costimulatory interactions. *J. Biol. Chem.* **271**, 26762–26771 (1996).
- An, H.-J. et al. Crystallographic and mutational analysis of the CD40–CD154 complex and its implications for receptor activation. *J. Biol. Chem.* **286**, 11226–11235 (2011).
- Singh, J. et al. The role of polar interactions in the molecular recognition of CD40L with its receptor CD40. *Protein Sci.* **7**, 1124–1135 (1998).
- Mempel, T. R., Henrickson, S. E. & Von Andrian, U. H. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature* **427**, 154–159 (2004).
- Swee, L. K., Lourido, S., Bell, G. W., Ingram, J. R. & Ploegh, H. L. One-step enzymatic modification of the cell surface redirects cellular cytotoxicity and parasite tropism. *ACS Chem. Biol.* **10**, 460–465 (2015).
- Eickhoff, S. et al. Robust anti-viral immunity requires multiple distinct T cell–dendritic cell interactions. *Cell* **162**, 1322–1337 (2015).
- Kretschmer, B., Kühl, S., Fleischer, B. & Breloer, M. Activated T cells induce rapid CD83 expression on B cells by engagement of CD40. *Immunol. Lett.* **136**, 221–227 (2011).
- Lesley, R., Kelly, L. M., Xu, Y. & Cyster, J. G. Naive CD4 T cells constitutively express CD40L and augment autoreactive B cell survival. *Proc. Natl Acad. Sci. USA* **103**, 10717–10722 (2006).
- Behrens, G. M. et al. Helper requirements for generation of effector CTL to islet β cell antigens. *J. Immunol.* **172**, 5420–5426 (2004).
- Liu, D. S., Loh, K. H., Lam, S. S., White, K. A. & Ting, A. Y. Imaging trans-cellular neurexin–neuroligin interactions by enzymatic probe ligation. *PLoS ONE* **8**, e52823 (2013).
- Slavoff, S. A., Liu, D. S., Cohen, J. D. & Ting, A. Y. Imaging protein–protein interactions inside living cells via interaction-dependent fluorophore ligation. *J. Am. Chem. Soc.* **133**, 19769–19776 (2011).
- Martell, J. D. et al. A split horseradish peroxidase for the detection of intercellular protein–protein interactions and sensitive visualization of synapses. *Nat. Biotechnol.* **34**, 774–780 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Ploegh for introducing us to sortase A; H. Yang, S. Markoulaki and R. Jaenisch for generating gene-targeted mice; A. Ting for NLG- and NRX-expressing constructs; and L. Mesin and C. F. Opel for technical advice. This work was funded by NIH grants DP5OD012146 and R01AI119006 to G.D.V. and a Starr Cancer Consortium grant to G.D.V. and N.H. G.P. was supported by the Swiss National Science Foundation Postdoctoral fellowship and the Cancer Research Institute Irvington Postdoctoral fellowship. G.V. is a Searle Scholar.

Author Contributions G.P. and G.D.V. conceived the study, designed and analysed experiments and wrote the manuscript. G.P. performed all experimental work (with the exception of gene-expression analysis), with sporadic assistance from A.Ch., J.M.T. and M.A. L.D.S., A.Cu. and N.H. contributed the gene-expression profiling work, including experiments and data analysis presented in Fig. 4g and Extended Data Fig. 10 and wrote the text for these experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to G.D.V. (victora@rockefeller.edu).

Reviewer Information Nature thanks M. Dustin, A. Esser-Kahn, T. Mempel and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Plasmids. All constructs were cloned into the pMP71 vector¹⁹, which was modified to express a fluorescent reporter (eGFP or Tomato) followed by the porcine teschovirus-1 self-cleavable 2A peptide²⁰ and the protein of interest. The SrtA sequence, including a terminal Flag-tag, was attached by a double 218 linker²¹ to the extracellular terminus of the modified receptor or ligand (C or N terminus, depending on protein topology). A five-glycine tag (G5) followed by a Myc tag was fused at the N terminus of modified receptors or ligands. The sequences of all constructs used are included in Supplementary Table 2.

Mice. C57BL/6J, CD45.1 (B6.SJL *Ptprc*^{ca}), *Cd40*^{-/-} (ref. 22), *Cd40lg*^{-/-} (ref. 23), *H2*^{-/-} (ref. 24), CD4-Cre-transgenic²⁵ and eCFP-transgenic²⁶ mice were purchased from The Jackson Laboratory (strain numbers 000664, 002014, 002928, 002770, 003584, 022071 and 004218, respectively). *Cd40*^{G5} and *Cd40lg*^{SrtA} mice were generated and maintained in our laboratories. B1-8^{hi} (ref. 27), J_HT (ref. 28) and OT-II TCR transgenic (Y chromosome)²⁹ mice were originally provided by M. Nussenzweig (Rockefeller University). All mice were housed in specific pathogen-free facilities at the Whitehead Institute for Biomedical Research and The Rockefeller University in accordance with institutional guidelines and ethical regulations. All protocols were approved by the Massachusetts Institute of Technology Committee for Animal Care and the Rockefeller University Institutional Animal Care and Use Committee. Male and female 5–12-week-old mice were used in all experiments.

Generation of *Cd40*^{G5} and *Cd40lg*^{SrtA} mice. The *Cd40*^{G5} mouse line was generated using CRISPR–Cas9 gene targeting by cytoplasmic injection of *Cas9* mRNA, chimeric single-guide RNA (sgRNA) and a repair oligonucleotide into fertilized C57BL/6 zygotes at the one-cell stage, as previously described^{30,31}.

The sequence for the dsDNA template for chimeric *Cd40*^{G5} sgRNA transcription was as follows (protospacer sequence is underlined): CGCTGTTAATACGACTCACTATAGGCTCTGTTT^{AGGTC}CTAGT^{TTT}TAGAGCTAGAAATAGCAAGTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGACCCGAGTCGGTGCTTTT.

The *Cd40*^{G5} repair oligonucleotide was synthesized as an ssDNA ultramer and PAGE-purified (Integrated DNA Technologies). The repair oligonucleotide sequence was as follows (differences from the wild-type C57BL/6 sequence are underlined): TGGCTGGCACAAATCACAGCACTGGCCATCGTGGAGGTACTGTTTGTCACTGCACGTAACGGTACCTCCTCCGCTCCACACTGCCCTAGATGTACCTAAAAACAGAGTGGACAGCTGGAAGGGATCTTCCACCGGC.

The *Cd40lg*^{SrtA} mouse line was generated using CRISPR–Cas9 gene targeting by cytoplasmic injection of *Cas9* mRNA, chimeric sgRNA, SCR7 (an NHEJ inhibitor, Excess Bioscience) and the repair plasmid into fertilized C57BL/6 zygotes at the one-cell stage, as described in ref. 32, with the exception that the final concentration of SCR7 used was 100 μM.

The sequence of the dsDNA template for chimeric *Cd40lg*^{SrtA} sgRNA transcription was as follows (protospacer sequence is underlined): CGCTGTTAATACGACTCACTATAGGAGAGTGGCTTCTCATCTTTGTTT^{AGGTC}CTAGTAAATAGCAAGTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGACCCGAGTCGGTGCTTTT.

The sequence of the *Cd40lg*^{SrtA} targeting construct is reported in Supplementary Table 2.

Cas9 mRNA was purchased from Sigma-Aldrich. Chimeric sgRNAs were *in vitro*-transcribed from a synthetic dsDNA template (gBlocks, Integrated DNA Technologies) using the MEGAshortscript T7 Transcription Kit (Thermo Fisher Scientific) and purified using Ampure XP beads (Beckman Coulter).

Isolation of splenic dendritic cells, CD4⁺ T cells and B cells. To isolate dendritic cells, spleens were collected, incubated for 30 min at 37 °C in RPMI, 2% FBS, 20 mM HEPES, 400 U ml⁻¹ type-IV collagenase (Worthington Biochemical) and disrupted to generate single-cell suspensions. Red-blood cells were lysed with ACK buffer (Lonza), and the resulting cell suspensions were filtered through a 70-μm mesh into PBS supplemented with 0.5% BSA and 2 mM EDTA (PBE). Dendritic cells were obtained by magnetic cell separation (MACS) using anti-CD11c beads (Miltenyi Biotec), as per the manufacturer's instructions. To isolate CD4⁺ T cells and B cells, spleens were processed as above, except for collagenase digestion, which was not performed. CD4⁺ T cells were isolated using the CD4⁺ T cell isolation kit (Miltenyi Biotec), whereas B cells were obtained by negative selection using anti-CD43 beads (Miltenyi Biotec), as per the manufacturer's instructions. To isolate Igλ⁺ B cells from B1-8^{hi} mice, B cells were stained with anti-Igλ-PE antibody and subsequently purified by negative selection using a combination of anti-CD43 and anti-PE magnetic beads (Miltenyi Biotec).

Cell transfers, immunizations and treatments. For dendritic cell transfer experiments, splenic dendritic cells isolated as described above were resuspended at 10⁷ cells per ml and treated with 10 μM OVA_{323–339} or LCMV-GP_{61–80} (both from Anaspec) in RPMI, 10% FBS, for 30 min at 37 °C. For cell labelling, CFSE was added to a final concentration of 2 μM during the last 5 min of incubation. Cells were then washed three times in RPMI, 10% FBS and resuspended at 2 × 10⁷ cells per ml in PBS supplemented with 0.4 μg ml⁻¹ LPS. Dendritic cells were injected (5 × 10⁵ cells in 25 μl) by subcutaneous injection into the hind footpad. For CD4⁺ T-cell transfer experiments, CD4⁺ T cells isolated as described above were resuspended at 3 × 10⁶ cells per ml in PBS and injected intravenously (3 × 10⁵ cells in 100 μl per mouse).

For immunization experiments, mice were immunized by subcutaneous injection into the hind footpad with 10 μg OVA adsorbed in alum (Imject Alum, Thermo Fisher Scientific) at 2:1 antigen:alum (v:v) ratio in 25 μl volume.

For LIPSTIC *in vivo* labelling experiments, biotin–LPETG (see below) was injected subcutaneously into the hind footpad (20 μl of 2.5 mM solution in PBS, equivalent to 50 nmol). Mice were injected six times 20 min apart, and popliteal lymph nodes were collected 40 min after the last injection. Mice were briefly anaesthetized with isoflurane at each injection.

For CD40L-blockade experiments *in vivo*, mice were injected intravenously with 200 μg of CD40L-blocking antibody (clone MR-1, BioXCell) at the indicated times. **Analysis of CD40L expression *in vivo*.** C57BL/6J dendritic cells were treated *ex vivo* with OVA_{323–339} and transferred subcutaneously (5 × 10⁵ per footpad) to *Cd40*^{-/-}, C57BL/6J or J_HT hosts. Eighteen hours later, 3 × 10⁵ *Cd40lg*^{SrtA/Y} OT-II or *Cd40lg*^{SrtA/Y} CD4-Cre OT-II CD4⁺ T cells were transferred intravenously and PLN were analysed 48 h after T-cell transfer.

Flow cytometry and cell sorting. Popliteal lymph nodes were collected, incubated for 30 min at 37 °C in RPMI, 2% FBS, 20 mM HEPES, 400 U ml⁻¹ type-IV collagenase (Worthington Biochemical), disrupted using disposable micropestles (Axygen) and filtered through a 70-μm cell strainer. Single-cell suspensions were washed with PBE, incubated at room temperature for 5 min with 1 μg ml⁻¹ of anti-CD16/32 (2.4G2, BioXCell) and then stained for cell surface markers at 4 °C for 15 min in PBE using the reagents listed in Supplementary Table 3. Cells were washed with PBS and stained with Zombie fixable viability dyes (Biolegend) at room temperature for 15 min and then fixed with Cytofix (BD Biosciences) before acquisition. In all *in vivo* experiments involving detection of biotin–LPETG SrtA substrate, an anti-biotin–PE antibody (Miltenyi Biotec) was exclusively used owing to its lower background compared to streptavidin conjugates. To eliminate unspecific signals derived from PE binding by a fraction of the B-cell population and thus reduce background, PE–Cy7 isotype control-positive cells were excluded from analysis. In all *in vivo* experiments involving detection of CD40L, a biotinylated anti-CD40L antibody (eBioscience) followed by an anti-biotin PE antibody (Miltenyi Biotec) was used. Samples were acquired on Fortessa or LSR-II flow cytometers (BD Biosciences) and data were analysed using FlowJo v.10.0.8 software.

RNA-sequencing of sorted dendritic cell populations. For the dendritic cell sorting experiment, *Cd40*^{G5/G5} dendritic cells were treated with OVA_{323–339} and transferred subcutaneously (5 × 10⁵ per footpad) into *Cd40*^{G5/G5} recipients. Eighteen hours later, 3 × 10⁵ *Cd40lg*^{SrtA/Y} CD4-Cre OT-II CD4⁺ T cells were transferred intravenously. Biotin–LPETG was administered subcutaneously (300 nmol per footpad) 46 h after T-cell transfer. Popliteal lymph nodes were processed 48 h after T-cell transfer and stained for surface markers as above and endogenous biotin⁺ and biotin⁺MHC-II^{hi}CD11c⁺CD11b⁺XCRI⁺ dendritic cells were sorted. As controls, MHC-II^{hi}CD11c⁺CD11b⁺XCRI⁺ dendritic cells were also sorted from *Cd40*^{-/-} mice treated as above, except that they received wild-type (instead of *Cd40*^{G5/G5}) dendritic cells and wild-type OT-II (instead of *Cd40lg*^{SrtA/Y} CD4-Cre OT-II) CD4⁺ T cells. Fresh cells were sorted (150 cells per sample) directly into plates containing TCL buffer (Qiagen) supplemented with 1% β-mercaptoethanol using a FACSAria II (BD Biosciences). RNA from sorted populations was isolated using Agencourt RNAClean XP beads (Beckman Coulter). Full-length cDNA and sequencing libraries were prepared using the Smart-seq2 protocol as previously described³³. Libraries were sequenced on a Nextseq500 (Illumina) to generate 38-base-pair, paired-end reads.

Raw sequencing data were processed as described³⁴. In brief, short sequencing reads were aligned to the UCSC mm10 transcriptome using Bowtie2 (v.2.1.0)³⁵. These alignments were used as input in RSEM (v.1.2.8)³⁶ to quantify gene expression levels for all UCSC mm10 genes in all samples. Data were normalized and analysed using the R software package DESeq2 (v.1.16.0)³⁷. Genes with low read counts, defined as those that do not have a normalized expression value greater than 100 in at least three samples, were filtered out, leaving 10,196 genes for the downstream analysis. The 500 genes with the largest variance were used for the principal component analysis and hierarchical clustering. For hierarchical clustering, the complete linkage clustering method was applied on pairwise

distances, defined as 1 minus the Pearson correlation coefficient. Paired differential expression analysis was performed for comparison between biotin⁺ and biotin⁻ dendritic cell samples. The differentially expressed genes were compared against the MSigDB database to compute for enrichment using the hypergeometric test³⁸.

Bone marrow chimaeras. C57BL/6J recipient mice were lethally irradiated with two doses of 450 Rads given 4 h apart. After irradiation, recipients were reconstituted by intravenous injection of haematopoietic cells collected from femurs and tibiae of donor mice. Mice were used for experiments 8–12 weeks after irradiation.

Western blot. Cells were lysed in sample buffer supplemented with 100 mM dithiothreitol. Cell lysates were heated at 98 °C for 5 min and then cleared by centrifugation at 15,000g for 10 min. Samples were separated by SDS-PAGE and transferred onto a nitrocellulose membrane. After blocking in 3% skim milk in PBS, membranes were incubated with 1–10 µg ml⁻¹ primary antibody in 3% skim milk in PBS overnight at 4 °C. After several washes in PBS and 0.1% Tween-20 (PBST), secondary antibodies coupled to HRP were applied in PBST for 1 h at room temperature when necessary. Blots were developed using Western Lightning ECL (Perkin-Elmer) and BioMax MR films (Kodak).

SrtA substrates. Biotin-aminohexanoic acid-LPETGS (C-terminal amide, 95% purity) was purchased from LifeTein (custom synthesis) and stock solutions prepared in PBS at 20 mM.

SELPETGG (C-terminal amide, 95% purity) was purchased from LifeTein (custom synthesis) and conjugated with AlexaFluor647 succinimidyl ester dye (Thermo Fisher Scientific). Reacted peptides were purified by HPLC.

Southern blot. Genomic DNA (10 µg) purified from mouse tails was digested with XbaI and separated on 0.8% agarose gel. Transfer and hybridization was performed as described³⁹. Blots were developed using Storage Phosphor Screens (GE Healthcare) and a Typhoon Imaging System (GE Healthcare). The sequence of the probe used was as follows: GGTCAACCTGGGTTCCTATAAATCTTG TCTTCCCCAAAAGGGGATAAATTCAGTAGACAGAGGCAGGTAGATCT CTGTGAGTCCCCAAGCTAGCTAGTCTGCATAACAAGTTGTAGGCCAGCT TCTGTTTCTTTCTCTCTCAAAAAAGAAAGCAGAAGTGAAGTGGGT AATGTATTTATTAACCTGAAAGAATCTGGTCTTTTCTCTCATTCAA ATGGTTCAAAAGTGAAACATCACAAACAAACATCCTTTATAGAGAA TTTGGGGTGCAATGTATCAG.

LIPSTIC in vitro. HEK293T cells (purchased from ATCC) were transfected using the calcium phosphate transfection kit (Thermo Fisher Scientific) with the indicated expression vectors. Forty hours after transfection, cells were detached using a non-enzymatic cell dissociation solution (Thermo Fisher Scientific), washed and resuspended at 10⁷ cell per ml in PBS. Cell populations transfected with G5- or SrtA-fusion constructs were mixed at a 1:1 ratio (10⁶ cells of each population) in a 1.5-ml conical tube, to which biotin-LPETG was added to a final concentration of 100 µM. Cells were incubated at room temperature for 30 min and washed three times with PBE to remove excess biotin-LPETG before FACS staining or western blot.

Imaging LIPSTIC in vitro. B cells and CD4⁺ T cells were isolated from mouse spleens as described above; B cells were activated with 25 µg ml⁻¹ LPS and 10 ng ml⁻¹ IL-4, whereas CD4⁺ T cells were activated with CD3/CD28 dynabeads and rat T-STIM conditioned medium (both from Thermo Fisher Scientific). Twenty-four hours later, cells were transduced with retroviral vectors. Transduced cells were sorted two days after transduction based on expression of the fluorescent reporter present in the retroviral vector. CD4⁺ T cells were incubated with AlexaFluor647-SELPETGG for 30 min at 37 °C, washed three times, and seeded together with B cells on 8-well Lab-Tek chamber slides (Sigma-Aldrich) previously coated with 12.5 µg ml⁻¹ ICAM (2 × 10⁵ cells per well, 1:1 ratio). Cells were immediately imaged using an Andor widefield microscope equipped with a live-cell incubation system. Images were acquired with a 40× objective every 45 s for 90 min using Metamorph software.

LIPSTIC ex vivo. Dendritic cells, B cells and CD4⁺ T cells were isolated from mouse spleens as described above.

Isolated dendritic cells were treated for 2.5 h at 37 °C with the indicated concentration of OVA_{323–339} or LCMV-GP_{61–80} peptides in RPMI, 10% FBS supplemented with LPS (10 µg ml⁻¹), washed three times and then seeded into U-bottom 96-well plates with purified CD4⁺ T cells (2 × 10⁵ cells per well, 1:1 ratio). Cells were co-cultured for 6 (Fig. 2 and Extended Data Fig. 5) or 24 h (Extended Data

Fig. 9), and biotin-LPETG was added at the indicated time of co-culture at a final concentration of 10 µM in complete medium. Blocking antibodies were added at the beginning of co-culture (Fig. 2) or at the indicated times (Extended Data Fig. 9) and used at a final concentration of 150 µg ml⁻¹.

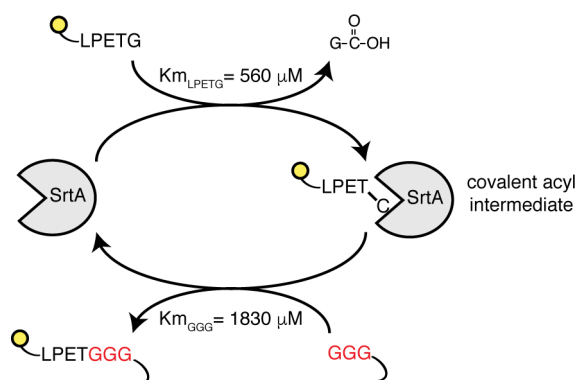
Purified B cells (either polyclonal or Igλ⁺ B1-8^{hi}) were cooled for 30 min on ice and then incubated for 45 min on ice with the indicated concentrations of NP-OVA (Biosearch Technologies). Cells were then washed twice and seeded into U-bottom 96-well plates with CD4⁺ T cells (2 × 10⁵ cells per well, 1:1 ratio) previously activated with CD3/CD28 dynabeads (Thermo Fisher Scientific) for 24 h. Cells were co-cultured for 18 h and biotin-LPETG was added during the last 30 min of co-culture at a final concentration of 100 µM in complete medium.

For all experiments, cells were washed three times with PBE before FACS staining to remove excess biotin-LPETG substrate.

Statistical analysis. Statistical tests were conducted using Prism (GraphPad) software. Gaussian distribution was confirmed by the Shapiro-Wilk normality test. Unpaired, two-tailed Student's *t*-tests and one-way ANOVA with Tukey's post hoc tests to further examine pairwise differences were used.

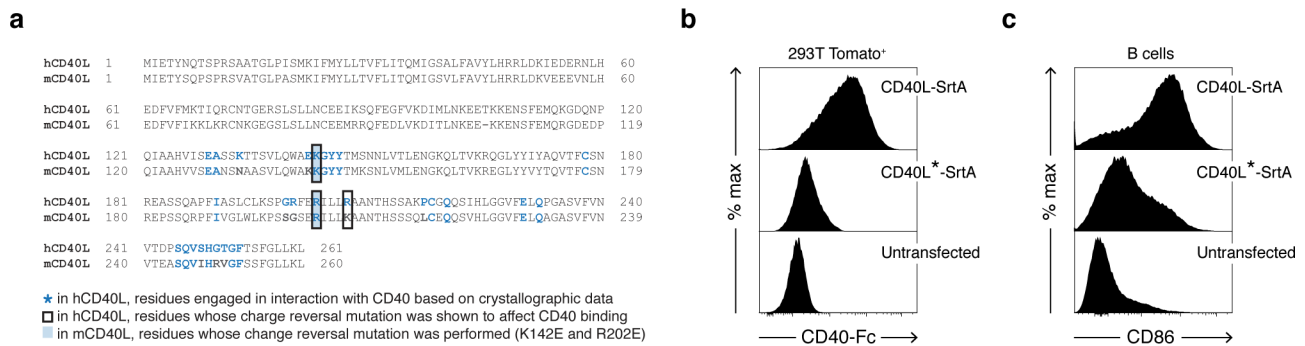
Data availability. RNA-sequencing data are deposited in GEO under accession number GSE107643. All other data are included within the article and the Supplementary Information or are available upon request from the corresponding author.

- Engels, B. *et al.* Retroviral vectors for high-level transgene expression in T lymphocytes. *Hum. Gene Ther.* **14**, 1155–1168 (2003).
- Kim, J. H. *et al.* High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS ONE* **6**, e18556 (2011).
- Whitlow, M. *et al.* An improved linker for single-chain Fv with reduced aggregation and enhanced proteolytic stability. *Protein Eng.* **6**, 989–995 (1993).
- Kawabe, T. *et al.* The immune responses in CD40-deficient mice: impaired immunoglobulin class switching and germinal center formation. *Immunity* **1**, 167–178 (1994).
- Renshaw, B. R. *et al.* Humoral immune responses in CD40 ligand-deficient mice. *J. Exp. Med.* **180**, 1889–1900 (1994).
- Madsen, L. *et al.* Mice lacking all conventional MHC class II genes. *Proc. Natl Acad. Sci. USA* **96**, 10338–10343 (1999).
- Lee, P. P. *et al.* A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity* **15**, 763–774 (2001).
- Hadjantonakis, A. K., Macmaster, S. & Nagy, A. Embryonic stem cells and mice expressing different GFP variants for multiple non-invasive reporter usage within a single animal. *BMC Biotechnol.* **2**, 11 (2002).
- Shih, T. A., Roederer, M. & Nussenzweig, M. C. Role of antigen receptor affinity in T cell-independent antibody responses *in vivo*. *Nat. Immunol.* **3**, 399–406 (2002).
- Gu, H., Zou, Y. R. & Rajewsky, K. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* **73**, 1155–1164 (1993).
- Barnden, M. J., Allison, J., Heath, W. R. & Carbone, F. R. Defective TCR expression in transgenic mice constructed using cDNA-based α- and β-chain genes under the control of heterologous regulatory elements. *Immunol. Cell Biol.* **76**, 34–40 (1998).
- Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**, 1370–1379 (2013).
- Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
- Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–542 (2015).
- Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
- Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–369 (2014).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
- Southern, E. Southern blotting. *Nat. Protoc.* **1**, 518–525 (2006).

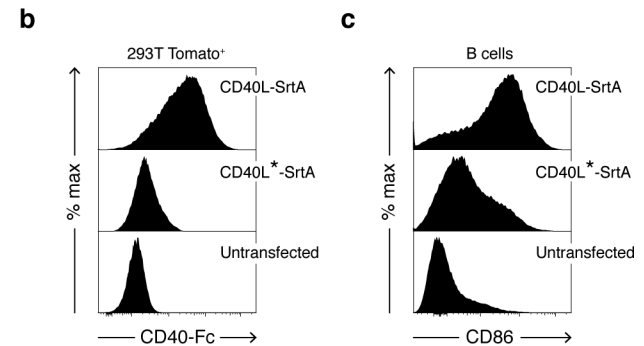


Extended Data Figure 1 | Schematic representation of the SrtA reaction.

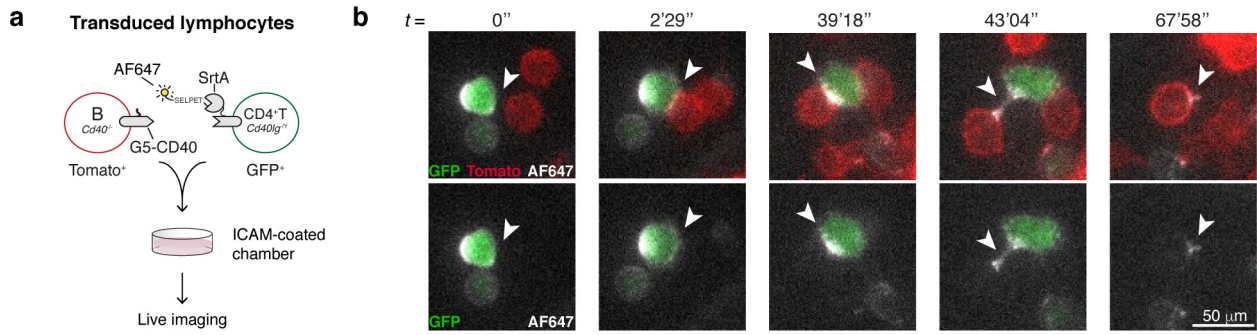
The SrtA enzyme recognizes the short amino acid sequence LPXTG (where X is any amino acid). Upon binding, SrtA forms a covalent acyl intermediate between the threonine of the substrate and the cysteine present in its catalytic pocket. The reaction proceeds with the formation of an amide bond between substrate threonine and an N-terminal glycine. Affinities displayed refer to engineered SrtA variants carrying P94S, D160N, and K196T mutations.



Extended Data Figure 2 | Two point mutations in the mouse CD40L coding sequence impair binding to CD40. **a**, Sequence alignment of human and mouse CD40L proteins. Owing to the lack of crystallographic data describing the mouse CD40–CD40L complex, we identified residues potentially engaged in CD40 binding on the basis of information available for the human CD40–CD40L complex. Residues in human CD40L sequences engaged in the interaction with CD40 based on crystallographic data are highlighted in blue. Among these, residues for which a charge reversal mutation was shown to affect CD40 binding are boxed. Filled boxes identify the residues in mouse CD40L for which a charge reversal mutation was performed (K142E and R202E). Mutations at equivalent locations in the human CD40L coding sequence (K143, R203) have also

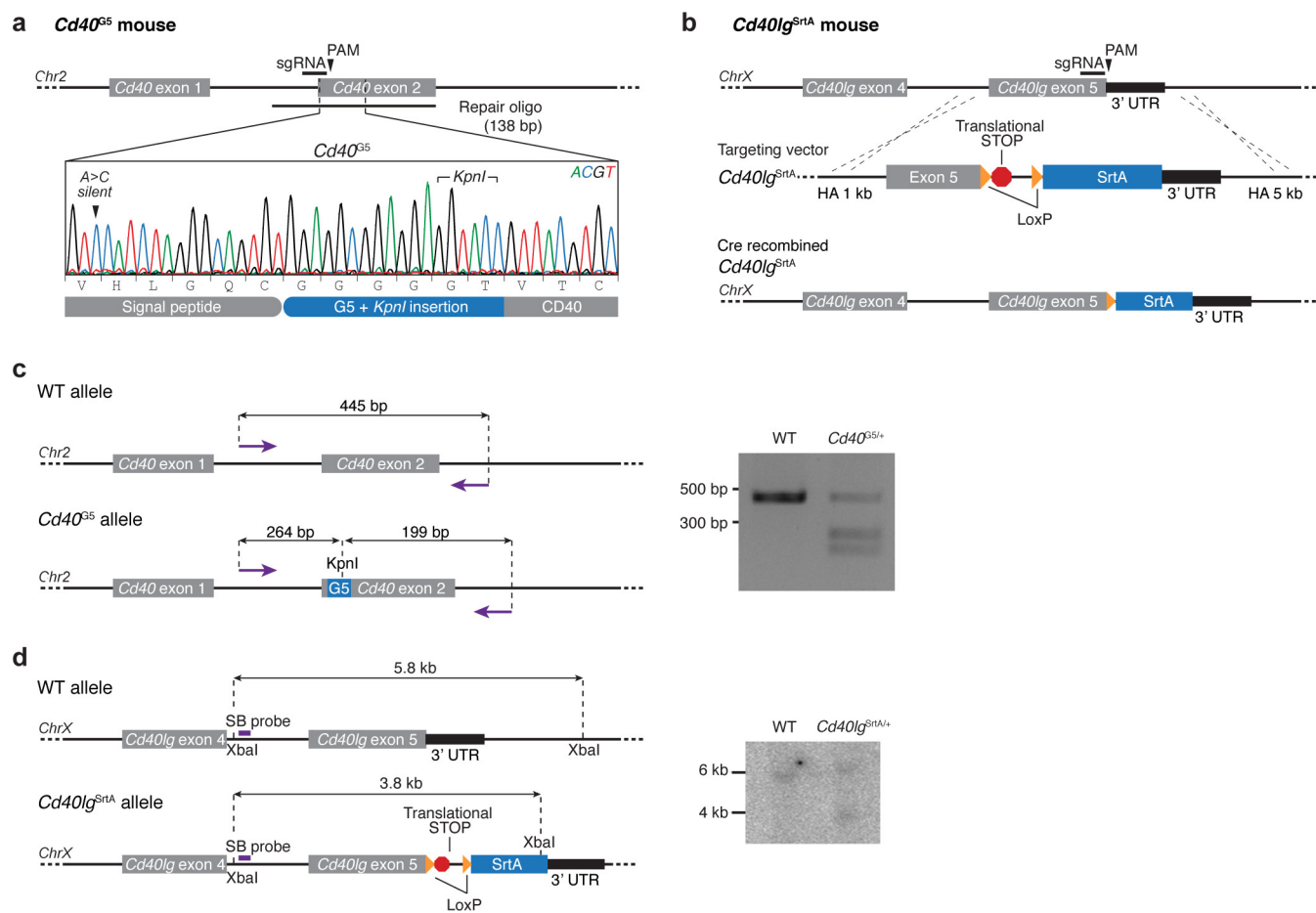


been detected in patients with hyper-IgM syndrome. CD40L with both mutations (K142E and R202E) is labelled as CD40L*. **b**, Binding of CD40 to CD40L–SrtA and CD40L*–SrtA. HEK293T cells were transfected with CD40L–SrtA or CD40L*–SrtA, incubated with CD40–Fc protein and analysed by flow cytometry. Histograms show severe impairment of CD40 binding to CD40L*–SrtA. **c**, B-cell activation by CD40L–SrtA and CD40L*–SrtA. Primary mouse B cells were cultured on a monolayer of HEK293T cells expressing CD40L–SrtA or CD40L*–SrtA. CD86 surface expression was analysed by flow cytometry 18 h later. Histograms show reduced upregulation of CD86 in B cells stimulated with CD40L*–SrtA. Data are representative of two independent experiments.



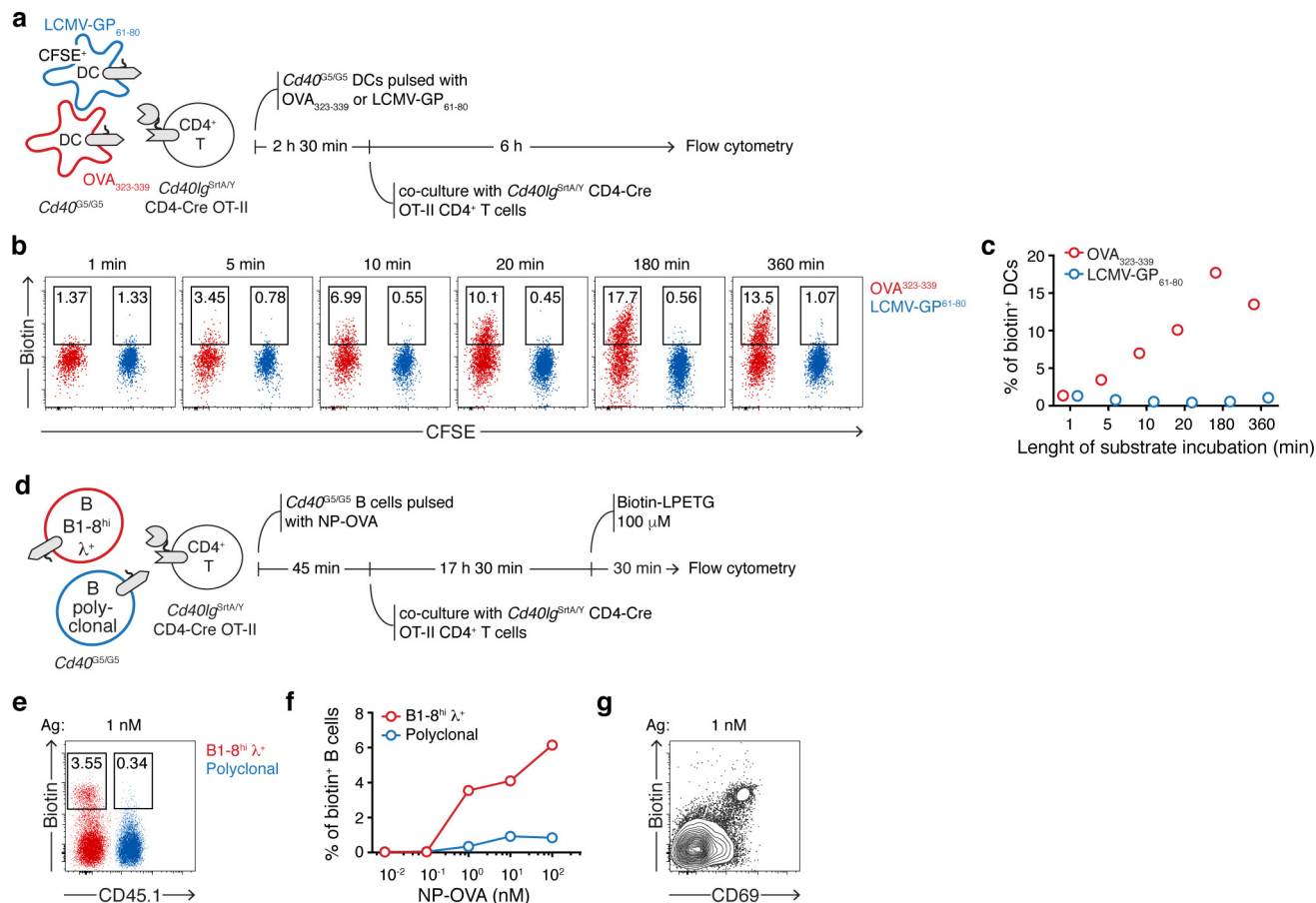
Extended Data Figure 3 | Imaging of LIPSTIC labelling. **a**, Experimental setup for live imaging of LIPSTIC labelling. *Cd40*^{-/-} B cells and *Cd40lg*^{-/-} CD4⁺ T cells were transduced with G5-CD40 (Tomato reporter) or CD40L-SrtA (GFP reporter), respectively. CD40L-SrtA⁺ T cells were loaded with AlexaFluor647-SELPETGG, mixed with G5-CD40⁺ B cells

on intercellular adhesion molecule (ICAM)-coated chambers to allow interactions and were immediately imaged. **b**, Time-series showing transfer of AlexaFluor647-SELPETGG (white) from CD40L-SrtA⁺ T cells (green) to G5-CD40⁺ B cells (red) upon interaction. Data are representative of two independent experiments.



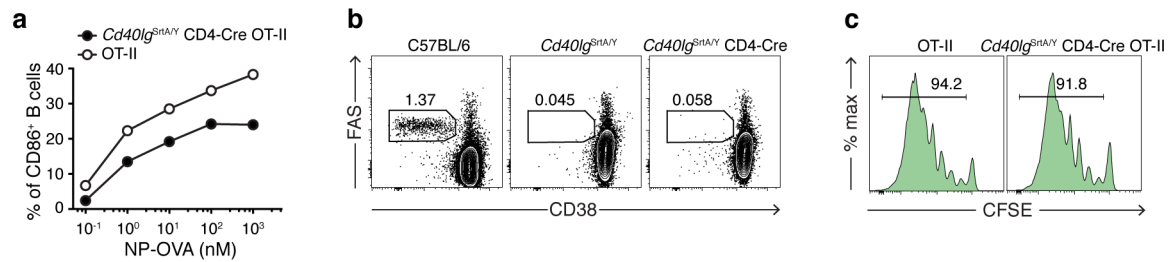
Extended Data Figure 4 | Generation of *Cd40^{G5}* and *Cd40lg^{SrtA}* gene-targeted mice. **a, b**, Schematic representation and CRISPR–Cas9 genome-editing strategy for the *Cd40^{G5}* (**a**) and *Cd40lg^{SrtA}* (**b**) alleles. HA, homology arm; PAM, protospacer adjacent motif. **c**, Restriction fragment length polymorphism analysis of *Cd40^{G5/+}* mice. PCR products generated using primers surrounding the G5 insertion site were

digested with KpnI and analysed by electrophoresis on an agarose gel. WT, wild type. Data are representative of at least two experiments. **d**, Southern blot analysis of a *Cd40lg^{SrtA/+}* mouse. Genomic DNA was extracted, digested with XbaI, and transferred onto a nitrocellulose membrane after electrophoresis on an agarose gel. Genomic DNA fragments were detected using a probe annealing between exons 4 and 5.



Extended Data Figure 5 | LIPSTIC labelling *ex vivo*. **a**, Experimental setup used in **b**, **c** to assess the influence of substrate incubation length on intercellular labelling between primary dendritic cells and CD4⁺ T cells. Cd40^{G5/G5} dendritic cells populations were separately treated with 1 μM of OVA₃₂₃₋₃₃₉ or LCMV-GP₆₁₋₈₀, mixed and co-cultured for 6 h with Cd40lg^{SrtA/Y} CD4-Cre OT-II CD4⁺ T cells. Biotin-LPETG was added during the final 1, 5, 10, 20, 180 min of co-culture or for the entire co-culture time (360 min) at a final concentration of 10 μM, and cells were analysed by flow cytometry. **b**, Flow cytometric analysis of co-cultured dendritic cells incubated with biotin-LPETG for the indicated times. **c**, Percentage of biotin⁺ dendritic cells gated as in **b**. **d**, Experimental setup used in **e**–**g** to analyse intercellular labelling *ex vivo* between primary B cells and CD4⁺ T cells. Two populations of Cd40^{G5/G5} B cells that either carried a wild-type polyclonal B-cell receptor repertoire or expressed the

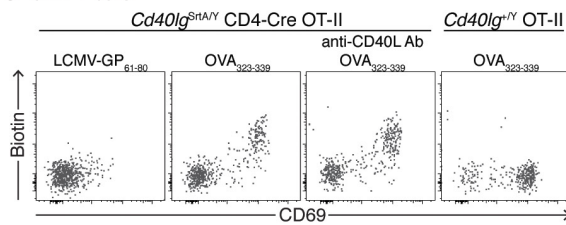
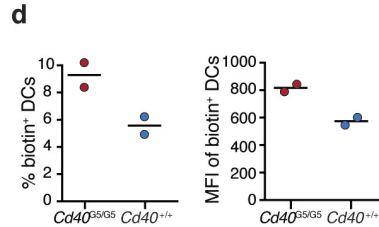
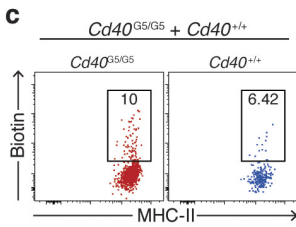
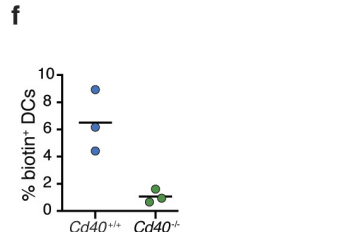
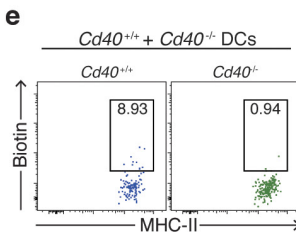
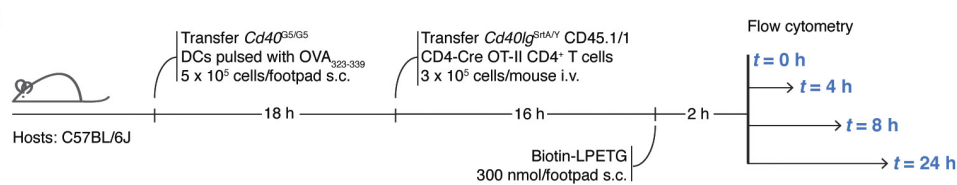
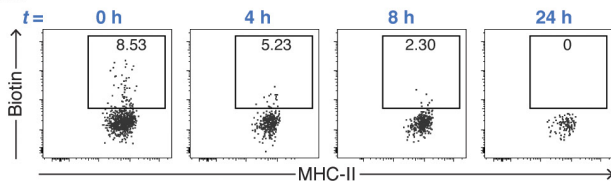
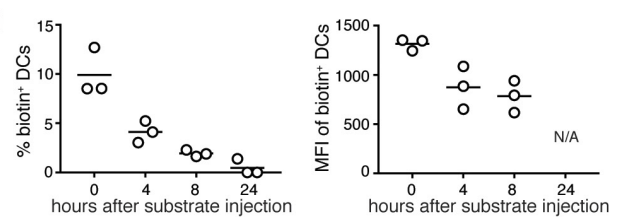
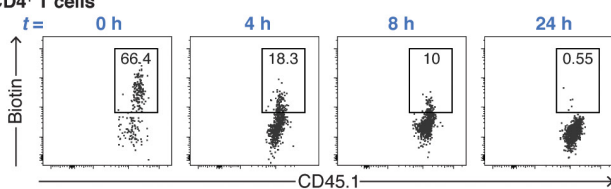
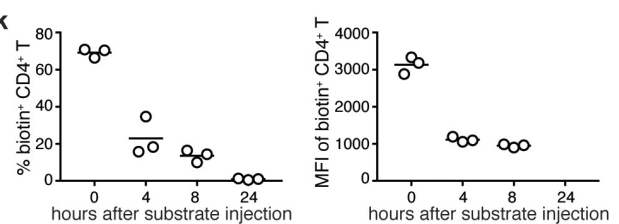
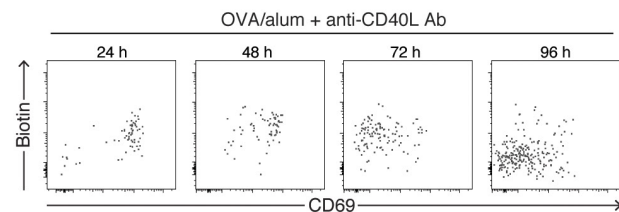
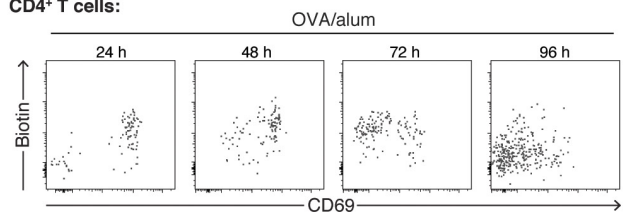
B1-8^{hi} Ig heavy chain, which when paired to an Igλ light chain confers specificity towards the hapten 4-hydroxy-3-nitrophenylacetyl (NP), were mixed and treated with the indicated concentrations of NP-OVA. Cells were then co-cultured for 18 h with Cd40lg^{SrtA/Y} CD4-Cre OT-II CD4⁺ T cells. Biotin-LPETG was added during the last 30 min of co-culture at a final concentration of 100 μM, and cells analysed by flow cytometry. **e**, Flow cytometric analysis of B cells treated with 1 nM of NP-OVA showing preferential biotin labelling of B1-8^{hi}λ⁺ B cells. **f**, Percentage of biotin⁺ B cells among polyclonal and B1-8^{hi}λ⁺ populations at the indicated NP-OVA concentrations. **g**, Flow cytometric analysis of B cells treated with 1 nM of NP-OVA showing positive correlation between biotin labelling and expression of the activation marker CD69. Data are representative of three independent experiments.



Extended Data Figure 6 | Characterization of *Cd40lg^{SrtA/Y}* T cells.

a, Upregulation of CD86 on B cells by CD40L–SrtA. B1-8^{hi} λ⁺ B cells were treated with the indicated concentrations of NP-OVA and co-cultured with *Cd40lg^{SrtA/Y}* CD4-Cre OT-II or wild-type OT-II CD4⁺ T cells for 18 h. Cells were analysed by flow cytometry. The percentage of CD86⁺ B cells when co-cultured with *Cd40lg^{SrtA/Y}* CD4-Cre OT-II or wild-type OT-II T cells in the presence of indicated concentrations of NP-OVA is shown. **b**, Germinal centre formation in *Cd40lg^{SrtA/Y}* and *Cd40lg^{SrtA/Y}* CD4-Cre mice. C57BL/6J, *Cd40lg^{+/Y}* and *Cd40lg^{SrtA/Y}* CD4-Cre mice were immunized subcutaneously with 20 μg of NP-OVA in alum at the base of the tail. Inguinal lymph nodes were analysed by flow cytometry 12 days after immunization. Dot plots show the absence of germinal centre formation in both *Cd40lg^{SrtA/Y}* and *Cd40lg^{SrtA/Y}* CD4-Cre mice, suggestive

of an impaired ability of *Cd40lg^{SrtA/Y}* T cells to activate B cells. A similar phenotype is observed regardless of the presence of Cre recombination, which is likely because of the addition of a translated LoxP site to the C terminus of the CD40L protein. **c**, *In vivo* expansion of *Cd40lg^{SrtA/Y}* CD4-Cre OT-II CD4⁺ T cells. 5×10^5 *Cd40^{G5/G5}* dendritic cells treated *ex vivo* with OVA_{323–339} were injected subcutaneously into the hind footpad of C57BL/6J recipients. After 18 h, 3×10^5 CFSE-labelled *Cd40lg^{SrtA/Y}* CD4-Cre OT-II (or wild-type OT-II as control) CD4⁺ T cells were transferred intravenously. PLNs were analysed by flow cytometry 72 h after T-cell transfer. Histograms show comparable expansion of both transferred T-cell populations, as indicated by CFSE dilution. Data are representative of two independent experiments.

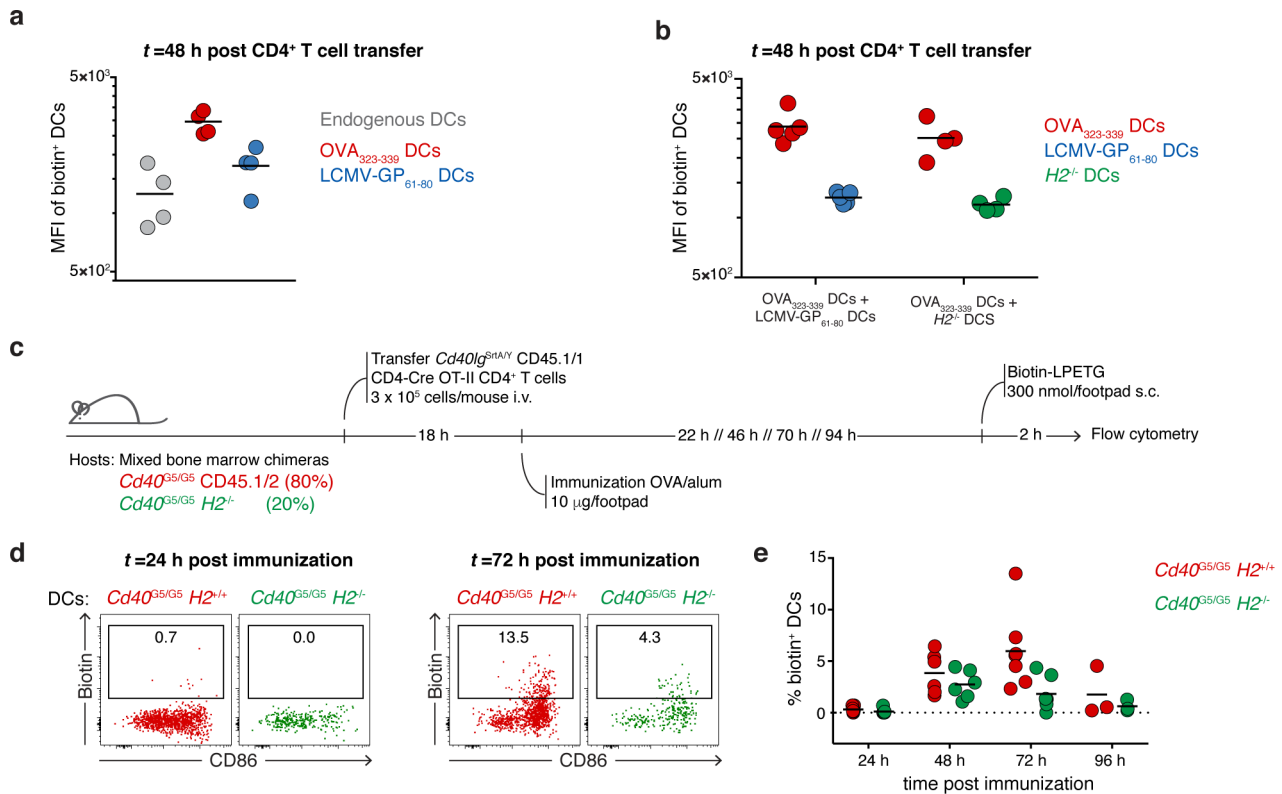
a CD4⁺ T cells:**c****e****g****h** DCs**i****j** CD4⁺ T cells**k****l** CD4⁺ T cells:

Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Characterization of LIPSTIC labelling *in vivo*.

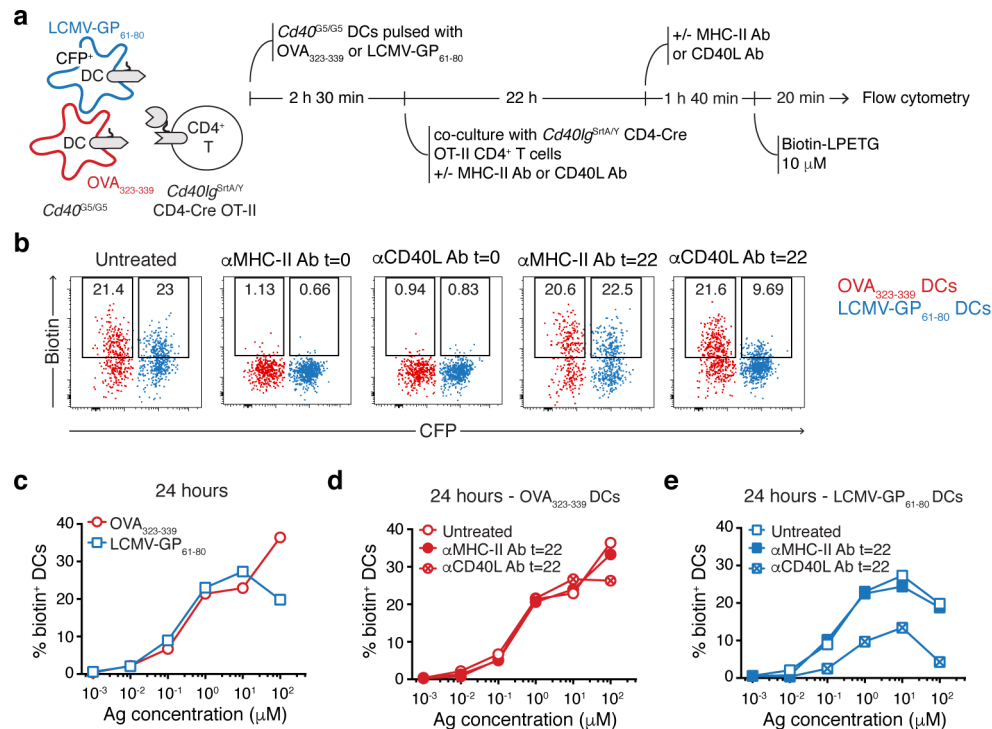
a, CD40L–SrtA expression in *Cd40lg^{SrtA/Y}* CD4–Cre OT-II CD4⁺ T cells *in vivo* after dendritic cell transfer. Mice were treated as in Fig. 3a. Flow cytometric analysis of PLN cells shows transferred *Cd40lg^{SrtA/Y}* CD4–Cre OT-II CD4⁺ T cells expressing CD40L–SrtA as revealed by the detection of biotin staining (formation of an acyl intermediate between SrtA and biotin–LPETG). CD40L–SrtA expression requires dendritic cell presentation of cognate antigen (OVA_{323–339}), is not affected by CD40L-blocking antibody treatment and positively correlates with the activation marker CD69. **b**, Detection background in major leukocyte populations. *Cd40^{G5/G5}* CD45.1/2 dendritic cells were treated with OVA_{323–339} and transferred subcutaneously (5×10^5 per footpad) into *Cd40^{G5/G5}* recipients. After 18 h, 3×10^5 *Cd40lg^{SrtA/Y}* CD45.1/1 CD4–Cre OT-II (or *Cd40lg^{+/Y}* CD45.1/1 OT-II lacking Cre expression as control) CD4⁺ T cells were transferred intravenously. Biotin–LPETG (or PBS as control) was administered subcutaneously (300 nmol per footpad) 10 to 12 h after T-cell transfer and PLN cells were analysed by flow cytometry. Plots show biotin staining among B cells, CD4⁺ T cells and dendritic cells. **c**, Efficiency of labelling of *Cd40^{G5/G5}* and *Cd40^{+/+}* dendritic cells after T cell–dendritic cell interaction *in vivo*. *Cd40^{G5/G5}* and *Cd40^{+/+}* dendritic cells were treated *ex vivo* with OVA_{323–339}, mixed and injected subcutaneously into C57BL/6J recipients (5×10^5 per footpad). After 18 h, 3×10^5 *Cd40lg^{SrtA/Y}* CD4–Cre OT-II CD4⁺ T cells were transferred intravenously. Biotin–LPETG was administered subcutaneously (300 nmol per footpad) 10–12 h after T-cell transfer. Dot plots show flow cytometric analysis of transferred *Cd40^{G5/G5}* and *Cd40^{+/+}* dendritic cells. **d**, Percentage (left) and MFI (right) of biotin⁺ dendritic cells (gated as in **c**) among transferred dendritic cell populations. Labelling of *Cd40^{+/+}* dendritic cells probably reflects biotin–LPETG transfer onto endogenous N-terminal glycines. Each symbol represents one mouse; bars indicate the mean. **e**, Labelling of

endogenous N-terminal glycines requires CD40L–CD40 interaction. Experimental setup as in **c**, except that a mixture of C57BL/6J and *Cd40^{−/−}* dendritic cells was transferred. Dot plots show flow cytometric analysis of transferred *Cd40^{+/+}* and *Cd40^{−/−}* dendritic cells. **f**, Percentage of biotin⁺ dendritic cells gated as in **e** among transferred dendritic cell populations. Each symbol represents one mouse; bars indicate the mean. **g**, Graphic representation of the experimental protocol used in **h–k** to determine the clearance of surface biotin labelling. *Cd40^{G5/G5}* dendritic cells were treated with OVA_{323–339} and transferred subcutaneously (5×10^5 per footpad) into C57BL/6J recipients. After 18 h, 3×10^5 *Cd40lg^{SrtA/Y}* CD4–Cre OT-II CD4⁺ T cells were transferred intravenously, biotin–LPETG was administered subcutaneously (300 nmol per footpad) 10–12 h after T-cell transfer. PLN cells were collected and analysed by flow cytometry 0, 4, 8 or 24 h after the final biotin–LPETG injection. **h**, Flow cytometric analysis of PLN cells showing biotin labelling of transferred *Cd40^{G5/G5}* dendritic cells at the indicated hours after biotin–LPETG administration. **i**, Percentage (left) and MFI (right) of biotin⁺ dendritic cells among transferred *Cd40^{G5/G5}* dendritic cells gated as in **h**. Each symbol represents one mouse; bars indicate the mean. **j**, Flow cytometric analysis of PLN cells showing biotin labelling of transferred *Cd40lg^{SrtA/Y}* CD4–Cre OT-II CD4⁺ T cells at the indicated time points after biotin–LPETG administration. **k**, Percentage (left) and MFI (right) of biotin⁺ cells among transferred *Cd40lg^{SrtA/Y}* CD4–Cre OT-II CD4⁺ T cells gated as in **h**. Each symbol represents one mouse; bars indicate the mean. **l**, CD40L–SrtA expression in *Cd40lg^{SrtA/Y}* CD4–Cre OT-II CD4⁺ T cells *in vivo* after immunization. Mice were treated as in Fig. 3d. Flow cytometric analysis of PLN cells showing transferred *Cd40lg^{SrtA/Y}* CD4–Cre OT-II CD4⁺ T cells in mice left untreated (left) or treated with CD40L-blocking antibody 4 h before PLN collection (right). Data are representative of two independent experiments.



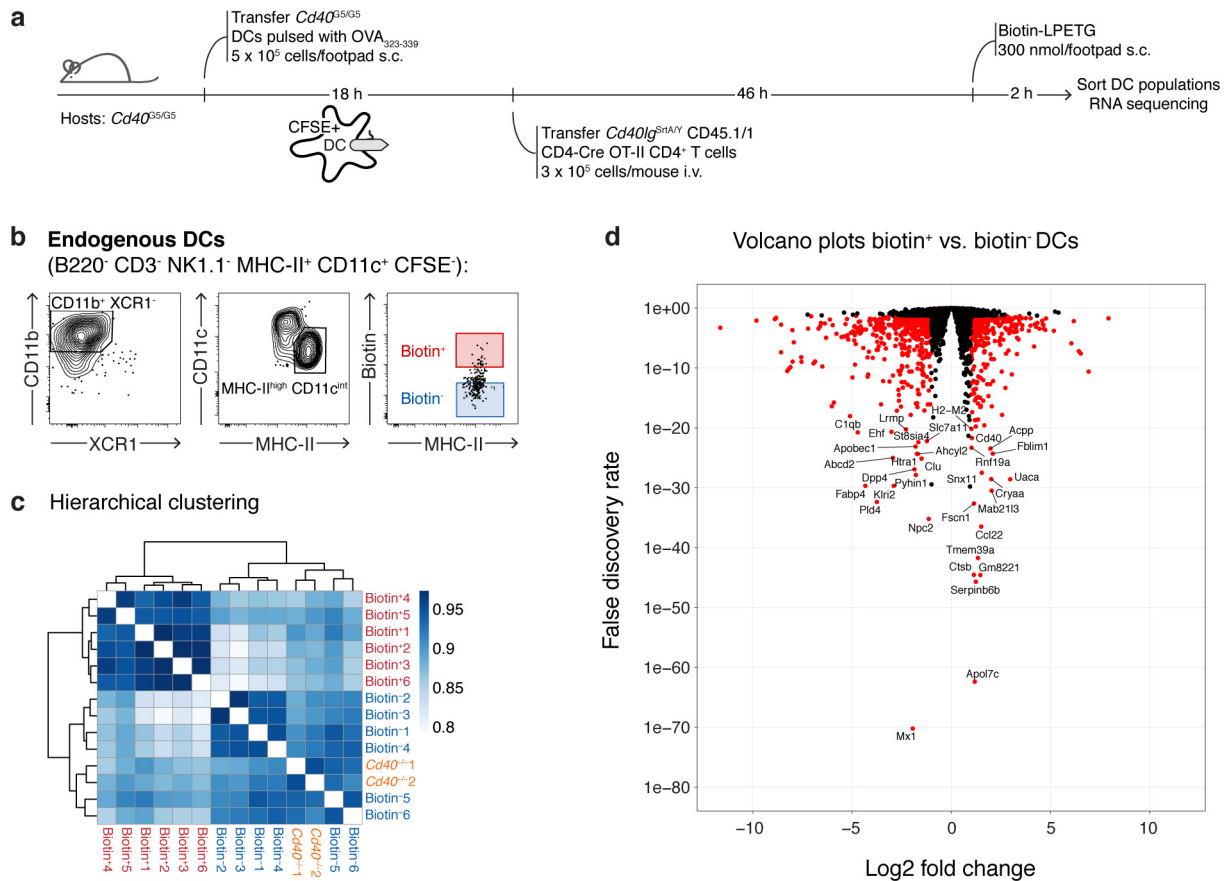
Extended Data Figure 8 | CD40–CD40L interaction between CD4⁺ T cells and dendritic cells *in vivo* can occur in an antigen-independent manner. **a**, MFI of biotin⁺ dendritic cells 48 h after T-cell transfer in mice treated as in Fig. 4a. Each symbol represents one mouse; bars indicate the mean. Data are pooled from two independent experiments. **b**, MFI of biotin⁺ dendritic cells 48 h after T-cell transfer in mice treated as in Fig. 4d. Each symbol represents one mouse; bars indicate the mean. **c**, Graphic representation of the experimental protocol used in **d**, **e**. C57BL/6J mice were lethally irradiated and reconstituted with a mixture of *Cd40*^{G5/G5} (80%) and *Cd40*^{G5/G5};H2^{L/-} (20%) bone marrow. After reconstitution, bone marrow chimaeras received 3×10^5 *Cd40lg*^{SrtA/Y}

CD4-Cre OT-II CD4⁺ T cells intravenously and were immunized the following day with 10 μ g of OVA in alum in the hind footpad. PLN were analysed 24, 48, 72 and 96 h after immunization. Biotin-LPETG was administered subcutaneously (300 nmol per footpad) during the last 2 h before analysis. **d**, Flow cytometric analysis of PLN cells showing biotin labelling of endogenous *Cd40*^{G5/G5} and *Cd40*^{G5/G5};H2^{L/-} dendritic cells at 24 or 72 h after immunization. **e**, Percentage of biotin⁺ dendritic cells among *Cd40*^{G5/G5} and *Cd40*^{G5/G5};H2^{L/-} dendritic cells gated as in **d**. Each symbol represents one mouse; bars indicate the mean. Data are pooled from two independent experiments.



Extended Data Figure 9 | CD40–CD40L interaction between CD4⁺ T cells and dendritic cells *ex vivo* can occur in an antigen-independent manner. **a**, Experimental setup used in **b–e**. Two *Cd40^{G5/G5}* dendritic cell populations were individually treated with the indicated concentrations of either OVA_{323–339} or LCMV-GP_{61–80}, mixed and co-cultured for 24 h with *Cd40lg^{SrtA/Y}* CD4-Cre OT-II CD4⁺ T cells. Biotin–LPETG was added during the last 20 min of co-culture at a final concentration of 10 μM, and

cells analysed by flow cytometry. Where indicated, CD40L- or MHC-II-blocking antibodies were added at a final concentration of 150 μg ml^{−1} either at the beginning of co-culture (*t* = 0) or 2 h before analysis (*t* = 22). **b**, Flow cytometric analysis of dendritic cells treated with 1 μM peptides showing biotin labelling. **c–e**, Percentage of biotin⁺ dendritic cells gated as in **b**. Data are representative of three independent experiments.



Extended Data Figure 10 | RNA sequencing analysis of sorted biotin⁺ dendritic cells. **a**, Graphic representation of the protocol for dendritic cell sorting. Five hundred thousand *Cd40^{G5/G5}* CFSE-labelled dendritic cells treated *ex vivo* with OVA₃₂₃₋₃₃₉ were injected subcutaneously into the hind footpad of *Cd40^{G5/G5}* recipients. After 18 h, 3 × 10⁵ *CD40lg^{SrtA/Y}* CD4-Cre OT-II CD4⁺ T cells were transferred intravenously. Biotin-LPETG was administered subcutaneously (300 nmol per footpad) during the last 2 h before analysis. PLNs were collected 48 h after T-cell transfer and dendritic cell populations were sorted by flow cytometry and later processed for RNA-sequencing analysis. As controls, dendritic cells were also sorted from *Cd40^{-/-}* mice, which were treated as above except that they received wild-type (instead of *Cd40^{G5/G5}*) dendritic cells and wild-type OT-II (instead of *CD40lg^{SrtA/Y}* CD4-Cre OT-II) CD4⁺

T cells. **b**, Gating strategy for sorting. Endogenous dendritic cells were first identified as B220⁻ CD3⁻ NK1.1⁻ MHC-II⁺ CD11c⁺ CFSE⁻. Sorting was restricted to CD11b⁺ XCR1⁻ dendritic cells showing an activated phenotype (MHC-II^{hi}), which represent the major population involved in bystander interactions. Biotin⁺ and biotin⁻ dendritic cells were gated as shown. **c**, Hierarchical clustering of transcriptomic profiles. Colour scheme is based upon Pearson correlation. Data are derived from a single experiment, *n* = 3. **d**, Volcano plots showing differential gene expression between biotin⁺ and biotin⁻ dendritic cells. All genes used for the differential expression analysis are shown; differentially expressed genes (log₂(fold change) > 1 and false-discovery rate < 0.05, see Methods) are coloured red. Data are derived from a single experiment, *n* = 3.

Structures of β -klotho reveal a 'zip code'-like mechanism for endocrine FGF signalling

Sangwon Lee¹, Jungyuen Choi¹, Jyotidarsini Mohanty¹, Leiliane P. Sousa¹, Francisco Tome¹, Els Pardon², Jan Steyaert², Mark A. Lemmon¹, Irit Lax¹ & Joseph Schlessinger¹

Canonical fibroblast growth factors (FGFs) activate FGF receptors (FGFRs) through paracrine or autocrine mechanisms in a process that requires cooperation with heparan sulfate proteoglycans, which function as co-receptors for FGFR activation^{1,2}. By contrast, endocrine FGFs (FGF19, FGF21 and FGF23) are circulating hormones that regulate critical metabolic processes in a variety of tissues^{3,4}. FGF19 regulates bile acid synthesis and lipogenesis, whereas FGF21 stimulates insulin sensitivity, energy expenditure and weight loss⁵. Endocrine FGFs signal through FGFRs in a manner that requires klothos, which are cell-surface proteins that possess tandem glycosidase domains^{3,4}. Here we describe the crystal structures of free and ligand-bound β -klotho extracellular regions that reveal the molecular mechanism that underlies the specificity of FGF21 towards β -klotho and demonstrate how the FGFR is activated in a klotho-dependent manner. β -Klotho serves as a primary 'zip code'-like receptor that acts as a targeting signal for FGF21, and FGFR functions as a catalytic subunit that mediates intracellular signalling. Our structures also show how the sugar-cutting enzyme glycosidase has evolved to become a specific receptor for hormones that regulate metabolic processes, including the lowering of blood sugar levels. Finally, we describe an agonistic variant of FGF21 with enhanced biological activity and present structural insights into the potential development of therapeutic agents for diseases linked to endocrine FGFs.

We used X-ray crystallography to determine the structure of the free and ligand-bound extracellular region of human β -klotho (sKLB) (Extended Data Fig. 1, Methods), in order to elucidate the mechanism of action of β -klotho in cell signalling via FGF21 stimulation. The overall structure of sKLB (2.2 Å resolution, Extended Data Table 1) features two tandem glycoside hydrolase-like domains, D1 (residues 53–507) and D2 (residues 521–968), which are connected by an unstructured and flexible linker (Fig. 1a). Each glycoside hydrolase-like domain can be recognized by multiple repeats of alternating layers of β -sheet and α -helix that define the (β/α)₈ fold (Extended Data Fig. 2a). The structure of KLB_{D1} (1.7 Å resolution, Extended Data Table 1) shown in Fig. 1b is virtually identical to the structure of D1 in the context of sKLB, with an overall C_α root mean square deviation (r.m.s.d.) value of 0.48 Å. Four loop regions in the structure of sKLB that contain potential N-glycosylation sites could not be modelled owing to poor electron density: a loop between H0 and S1 (residues 63–73; H denotes α -helix and S denotes β -sheet), a loop between H1b and H1c (residues 119–125), a loop between S9 and H9a (residues 538–574) and the C terminus of the protein (residues 968–983) (Extended Data Fig. 2a). With the exception of the C terminus, these loops are depicted in the sKLB structures as dashed lines (Fig. 1a).

Superimposing the structure of human cytosolic β -glucosidase (RCSB Protein Data Bank (PDB) code: 2ZOX) on the structures of each of the two glycoside hydrolase-like domains in sKLB gave C_α r.m.s.d. values of 1.08 Å for D1 and 1.39 Å for D2, respectively, which

demonstrates the strong similarity of both D1 and D2 to glycoside hydrolase family-1 (GH1) enzymes (Fig. 1d, e). GH1 enzymes hydrolyse glycosidic linkages between carbohydrate moieties (<http://www.cazy.org/GH1.html>) through a double-replacement mechanism mediated by two conserved glutamate residues located in their active sites⁶. In each of the sKLB domains, one of these two 'catalytic' glutamates is replaced by another amino acid (Fig. 1d–f): the first glutamate in D1 is replaced by Asn241, and the second glutamate in D2 is replaced by Ala889. This indicates that neither glycoside hydrolase-like domain in β -klotho can function as an active glycoside-hydrolase enzyme. Structural alignment using the Dali server⁷ indicates that GH1 and GH5 members exhibit high structural similarities to each of the glycoside hydrolase-like domains of sKLB, suggesting a common evolutionary origin. Although the overall structures of the glycoside hydrolase-like domains in sKLB are very similar to GH1 enzymes, the two sKLB glycoside hydrolase-like domains exhibit important structural features that set them apart from GH1 enzymes.

The pocket in D1 that corresponds to the substrate-binding region in GH1 enzymes is largely occluded by a short helix, H6a (Fig. 1d and Extended Data Fig. 3a). Moreover, a helix-turn-strand element (H6a-turn-S6b) in this region, specific to β -klotho D1 (green in Fig. 1d), provides part of the FGF21-binding site (see below) and is quite distinct from the strand-helix-strand element in the corresponding regions of cytosolic β -glucosidase (shown in grey in Fig. 1d). Other features unique to β -klotho include a short helix, H0 (Fig. 1d and Extended Data Fig. 3b), which begins with the first amino acid that follows the sKLB signal sequence (Phe53). This helix interacts with H5a, H6b and S5b, mostly through hydrophobic interactions, and precedes a disordered loop that is followed by the core structural elements of the (β/α)₈ fold. Glu416, the remaining catalytic residue in D1, is located at the bottom of the substrate-binding pocket (Fig. 1d and Extended Data Fig. 3a), and the orientation of the side chain of Glu416 is identical to the orientation of the side chain of the corresponding nucleophilic Glu373 residue of human cytosolic β -glucosidase.

The pocket in D2 that corresponds to the substrate-binding pocket in GH1 enzymes is not occluded by an α -helix in the D2 domain, but is instead accessible and occupied by a 2-(N-morpholino)ethanesulfonic acid (MES) molecule from the crystallization buffers (Fig. 1c). The morpholine ring of MES interacts with aromatic rings from three phenylalanines, Phe931, Phe826 and Phe942 (Fig. 1c), which also have a role in the interaction of sKLB with its ligands (see below). The D2 pocket is accessible in part because of the existence of a disordered region between S9 and H9a (Extended Data Fig. 2a), which produces a groove-like feature in this domain instead of the pocket that accommodates the substrate in the active site of GH1 members. The amino acid sequence and the length of this region vary considerably among GH1 members. The inter-domain interface of sKLB comprises an extensive network of both hydrophobic and polar interactions (Extended Data Fig. 3c) that encompasses a buried surface area⁸ of about 680 Å².

¹Department of Pharmacology and Yale Cancer Biology Institute, Yale School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520, USA. ²VIB Center for Structural Biology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium.

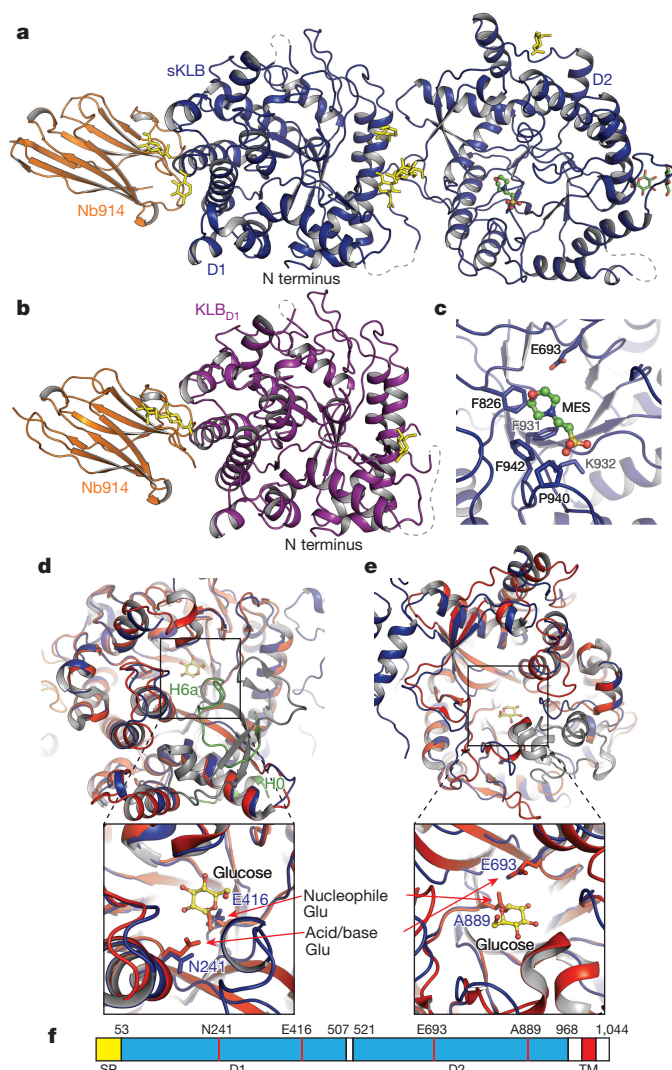


Figure 1 | Crystal structure of extracellular domain of β -klotho.

a, b, Structures of sKLB (**a**) (blue) and KLB_{D1} (**b**) (purple) in complex with nanobody Nb914 (orange) are shown as ribbon representations. Yellow sticks denote glycans attached to asparagine side chains; green sticks denote glucose molecules; the MES molecule is shown as a ball-and-stick representation; dashed grey lines indicate regions that do not show significant electron density. **c**, Side-chain atoms of amino acids in sKLB interacting with the MES molecule are shown as sticks. Also indicated is the location of Glu693, which is approximately 6 Å away from the bound MES molecule. **d, e**, The structure of human cytosolic β -glucosidase (red; PDB code: 2ZOX) is superimposed with that of D1 (**d**) and D2 (**e**) of sKLB (blue) with overall C_{α} r.m.s.d. values of 1.08 Å and 1.39 Å, respectively. Regions in sKLB that are different from β -glucosidase are coloured in green; regions in β -glucosidase that are different from sKLB are coloured in grey. A glucose molecule bound to β -glucosidase is shown as a ball-and-stick representation in yellow. Superimposition of D1 and D2 reveals locations of pseudo-catalytic glutamates. Note that one of the two conserved glutamates from each of the sKLB domains is replaced by an asparagine (for D1) or an alanine (for D2). **f**, Diagram of β -klotho that highlights the locations of the residues corresponding to the conserved glutamates in D1 and D2 of β -klotho. SP, signal peptide, TM, transmembrane.

Next, we determined the structure of sKLB in complex with C-terminal tail of FGF21 (FGF21_{CT}) at 2.6 Å resolution (Fig. 2a and Extended Data Table 1). Our final model contains amino acids Pro186–Ser209 from FGF21_{CT} bound to sKLB (Extended Data Fig. 4a–c), and exhibited clear electron density for FGF21_{CT} that lay across the middle of sKLB (Fig. 2b). FGF21_{CT} binds to an elongated

interface that spans D1 and D2 of sKLB; this binding has no influence on the structure of either individual domain, as judged by a C_{α} r.m.s.d. values of 0.33 Å for D1 and 0.49 Å for D2 when overlaid on the unoccupied sKLB structure. We observed a small change of 6° in the inter-domain angle⁹ when FGF21_{CT} bound to sKLB (Extended Data Fig. 4d). The FGF21_{CT}-binding region on sKLB is located on the opposite side of the molecule from the linker that connects D1 and D2. The flexibility of the linker may contribute to the inter-domain dynamic properties that enable complex formation with ligands and FGFRs. The sKLB–FGF21_{CT} structure shows two distinct binding sites for two different regions of the peptide: site 1 is located on D1 and site 2 is located in D2, with a distance of 30 Å between them.

Site 1 on sKLB D1 engages amino acids Pro186–Val197 of FGF21_{CT}, primarily through hydrophobic interactions (Fig. 2e, Extended Data Fig. 4a, b). Site 1 involves a surface created on D1 by H6a, H7, the loop between S6b and H6b, and the loop between S7 and H7. Most notably, the region of the bound peptide ligand that associates with site 1 adopts an unusually compact and rigid structure through the formation of several well-defined turns (Fig. 2e), as follows: (1) Asp187–Val188–Gly189–Ser190 form a type I β -turn (shown in orange in Fig. 2e) through hydrogen bonding of the carboxyl oxygen of Asp187 with the backbone nitrogen of Gly189, and of the backbone carbonyl of Asp176 with the backbone amide of Ser190; (2) Ser190–Ser191–Asp192 form an ST turn (a structural feature containing hydrogen-bonded serine or threonine residues, shown in yellow in Fig. 2e) through hydrogen bonding of the Ser190 hydroxyl with the backbone amide of Asp192; (3) Asp192–Pro193–Leu194–Ser195 (shown in light blue in Fig. 2e) form a type I β -turn (or an Asx turn that resembles a Schellman loop) through hydrogen bonding of the side-chain carboxyl of Asp192 with the Met196 and Val197 backbone amides, and of the Asp192 backbone carbonyl with the backbone amide of Ser195. These consecutive turns also support a long-range hydrogen bond between the Asp187 backbone amide and the Pro193 carbonyl. These intramolecular interactions cooperate to form a well-defined structural element that makes multiple specific contacts with sKLB, burying a relatively large surface area of 606 Å².

Site 2 interactions with FGF21_{CT} contrast markedly with site 1 interactions, and comprise a network of intermolecular interactions of the sort typically observed between proteins and short peptides (Fig. 2f and Extended Data Fig. 4a, b). Residues 200–209 of the FGF21_{CT} peptide project into what would be the substrate-binding site occupied by glycosides that D2 of sKLB would hydrolyse if it were an active GH1 enzyme (Fig. 3 and see later). It is also noteworthy that half of the sequence of this part of FGF21_{CT} (S-Q-G-R-S-P-S-Y-A-S) consists of residues with side-chain hydroxyl groups, suggesting that this region of FGF21 may indeed mimic a glycoside substrate. Given these characteristics, a notable feature of site 2 is the interaction between the side-chain carboxyl group of Glu693 in sKLB and hydroxyl groups of Ser204 and Ser206 in FGF21_{CT} (Fig. 3d). Glu693 corresponds to one of the two conserved catalytic glutamates, and would function as a general acid–base catalyst in the Koshland double-displacement reaction of glycoside hydrolases; by contrast, in D2 the potential nucleophilic glutamate is replaced by alanine.

Amino acids 198–200 of FGF21_{CT}, which connect ligand-binding sites 1 and 2, do not make substantial contacts with sKLB. In addition, the electron densities in omit maps (Fig. 2b) and B-factors (Fig. 2c) suggest that this region is flexible. This conclusion is consistent with the previous identification of an enzyme that cleaves FGF21 in this region, and which is known to abolish the binding of FGF21 to β -klotho^{10–15}. As this region of FGF21 is flexible and potentially accessible for proteolysis, cleavage between the binding regions of sites 1 and 2 could represent a mechanism for the termination of FGF21 signalling by targeted proteolysis.

The crystal structure of sKLB bound to FGF21_{CT} reveals how the basic framework of a glycoside hydrolase has evolved to become a specific receptor for endocrine FGFs. The β -glucosidase family

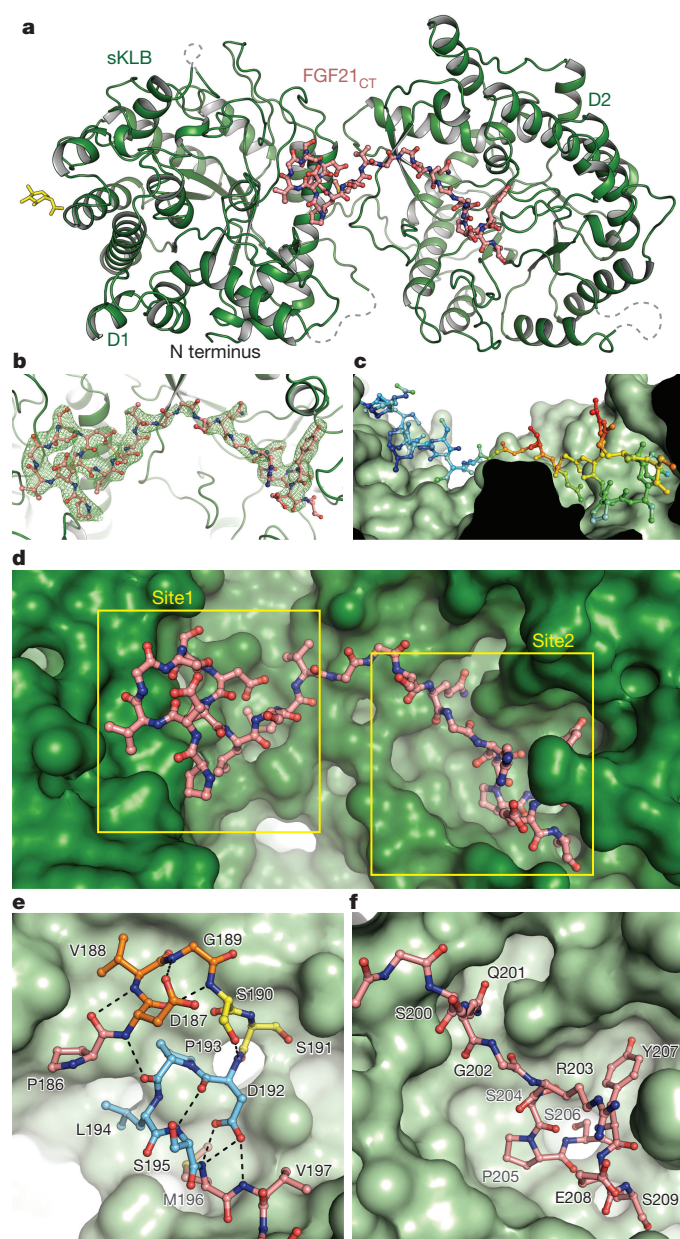


Figure 2 | Crystal structure of sKLB bound to FGF21_{CT} reveals two distinct binding sites. **a**, The structure of sKLB (green) in complex with FGF21_{CT} (salmon) is shown as a ribbon and ball-and-stick representation. Yellow sticks denote N-linked glycans. Nb914 is omitted for clarity. Grey dashed lines denote regions that do not exhibit significant electron densities. **b**, FGF21_{CT} binding site showing $|F_o| - |F_c|$ omit map contoured at 3.0σ for FGF21_{CT}. **c**, Surface of sKLB interacting with FGF21_{CT} are colour-coded according to B-factor values, which range from 52.76 \AA^2 (blue) to 103.63 \AA^2 (red). **d**, Surface representation of sKLB (green) highlighting two binding sites, site 1 and site 2 of FGF21_{CT} (salmon, ball-and-stick). **e**, Site 1 forms a series of internal hydrogen bonds (black dashed lines) through three consecutive turns (orange, yellow and light blue), creating a structural element that binds to D1 of sKLB. **f**, Site 2 interacts with the pseudo-substrate-binding region of D2 of sKLB.

of glycoside hydrolases catalyse the hydrolysis of disaccharides as well as longer oligosaccharides, and several crystal structures of β -glucosidases in complex with oligosaccharide substrates such as cellotetraose (*Paenibacillus polymyxa* BglB, PDB code: 2Z1S) and cellopentaose (*Oryza sativa* Bglu1, PDB code: 3F5K) have previously been determined^{16,17}. Superimposition of the crystal structures of substrate-bound β -glucosidases with the structure of sKLB in complex with FGF21_{CT} shows that the backbone of residues 200–209 from

FGF21_{CT} aligns very well with the location of oligosaccharides that occupy the catalytic pocket of β -glucosidases (Fig. 3a–c). The mode of interaction between the hydroxyls of Ser204 and Ser206 from FGF21_{CT} and the conserved glutamate in D2 of sKLB, together with hydrophobic interactions involving Pro205, are highly reminiscent of the substrate interactions seen for the glycoside hydrolases⁶, suggesting that the sKLB–FGF21_{CT} site 2 interaction is a pseudo-substrate interaction (Fig. 3d). Oligosaccharide substrates bound to the catalytic glutamic acid in the active sites of β -glucosidases lie in precisely the same position as the Ser204–Pro205–Ser206 motif of FGF21 bound to site 2 of sKLB. In addition, the residues in sKLB that form hydrophobic interactions with the Pro205 of FGF21—that is, Phe826, Phe931 and Phe942—align closely with the corresponding hydrophobic residues in β -glucosidases. These unexpected similarities indicate that the substrate-binding region of glycoside hydrolases evolved to recognize a sugar-mimicking Ser–Pro–Ser motif in FGF21 (Fig. 3e). As FGF19 also binds specifically to β -klotho, it is not surprising that FGF19 also contains a Ser211–Pro212–Ser213 motif at its C terminus (Extended Data Fig. 5), whereas FGF23—which does not bind to β -klotho—has no such sequence. Future studies that investigate how FGF23 recognizes α -klotho should provide guidance for the development of new treatments of metabolic disorders caused by impaired phosphate homeostasis¹⁸, as well as information on the unique evolutionary pathway that this family of proteins may have taken.

We next analysed the binding affinities between sKLB and FGF21, the wild-type FGF21 C terminus or a range of mutations of the C-terminal tail of FGF21 to investigate the contributions of different amino acids in FGF21 that take part in the interface of the ligand-occupied sKLB structure. We also investigated the effects of mutations in the two FGF21-binding sites of β -klotho on the ability of FGF21 to stimulate FGFR1 activation in transfected L6 rat myoblasts. These experiments validated the ligand-binding interfaces identified in the occupied sKLB structure and demonstrated that FGF21_{CT} binds in a cooperative manner to both site 1 and site 2 in β -klotho (for full details of these experiments, see Extended Data Figs 6, 7 and Supplementary Discussion).

Because endocrine FGFs have important roles in the control of metabolic processes, a variety of approaches have previously been used to develop therapeutic variants of these proteins^{12,19–24}. We reasoned that it should be possible to enhance the potency of FGF21, by introducing into its C-terminal tail mutations that strengthen interactions with β -klotho. We introduced a Leu194Phe mutation to increase hydrophobic interactions with neighbouring amino acids in site 1 of β -klotho, and an Arg203Trp mutation to replace cation– π interactions between Arg203 in FGF21 and His646 in site 2 of β -klotho with π – π interactions. We found that FGF21 with the Arg203Trp and Leu194Phe substitutions (FGF21_{WF}) bound to sKLB over tenfold more tightly than wild-type FGF21, with a dissociation constant (K_d) value of $3.4 \pm 1.2 \text{ nM}$ (Fig. 4a), and that the FGF21_{WF} mutant had an enhanced ability to stimulate FGFR1c autophosphorylation and MAP kinase stimulation in L6 cells that co-express β -klotho and FGFR1c (Fig. 4b and Extended Data Fig. 8).

These experiments show that, rather than serving as an alternative co-receptor for FGFR1c activation by endocrine FGFs, β -klotho functions as the primary high-affinity receptor for FGF21. We show that klotho proteins function as specific zip-code-like signals for targeting FGF21 (or two other endocrine FGFs) to cells and tissues, where they mediate their cellular responses by activating members of the FGFR families. The scheme presented in Fig. 4c depicts a model of how FGF21 binding to β -klotho enables it to activate a β -klotho–FGFR complex to promote cell signalling. In the model, FGFR1c and β -klotho monomers exist in equilibrium with β -klotho–FGFR heterodimers in the membrane. With a K_d value of approximately $1 \mu\text{M}$ for the binding of the FGFR1c extracellular region to sKLB (Extended Data Fig. 6b), a substantial portion of FGFR1c and β -klotho will be associated with one another at levels of around 10,000 copies per cell. FGF21 binds

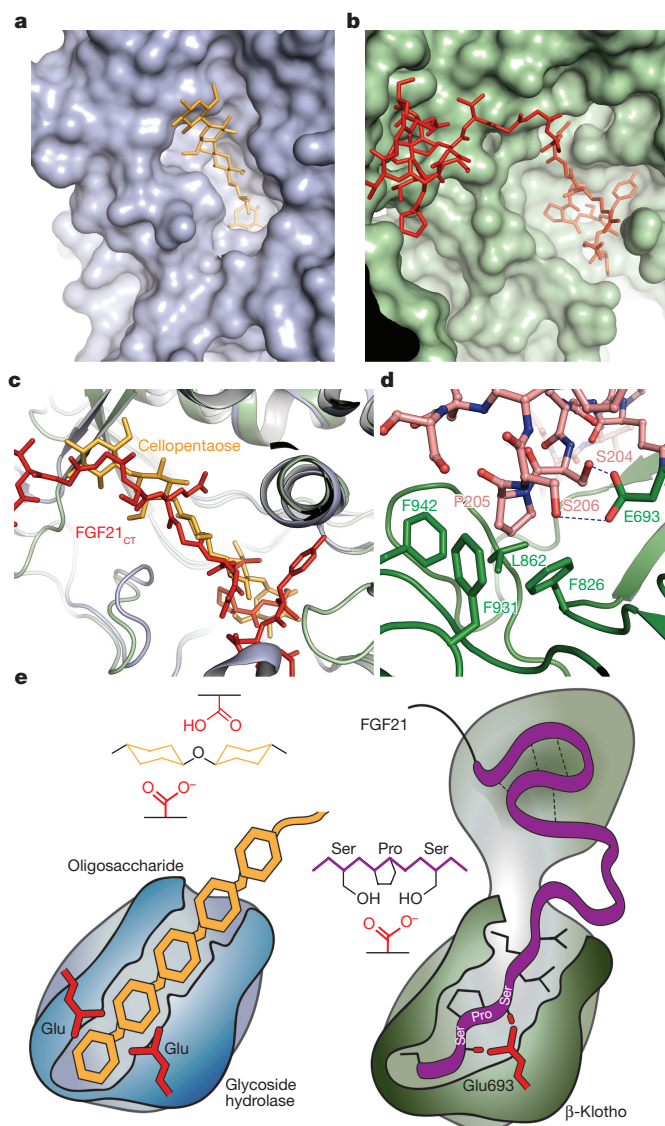


Figure 3 | Comparison of β -glucosidase and β -klotho structures and the evolution of a sugar-cutting enzyme into a receptor for endocrine FGF. **a, b**, The structure of rice β -glucosidase (**a**) (light blue, surface representation) in complex with cellopentaose (orange, stick representation) (PDB code: 3F5K) and site 2 of sKLB (**b**) (pale green, surface representation) in complex with FGF21_{CT} (red, stick representation). Cellopentaose binds to the active site of β -glucosidase and FGF21_{CT} binds to the corresponding pseudo-substrate binding site of β -klotho. **c**, Superimposition of the structures of cellopentaose-bound rice β -glucosidase and FGF21_{CT}-bound sKLB. **d**, Glu693 of β -klotho makes contacts with Ser-Pro-Ser motif of FGF21 via interaction with the hydroxyl moieties of serine residues, mimicking the sugar hydroxyls in their interaction with glutamate residues in the catalytic site of β -glucosidase. **e**, Schematic diagram comparing the substrate-binding pocket including the two glutamic acid residues required for glycoside hydrolase activity with the ligand-binding pocket of β -klotho depicting interactions between Glu693 and the Ser-Pro-Ser motif.

with high affinity ($K_d = 43.5$ nM, Extended Data Fig. 6a) either to β -klotho monomers or to pre-existing β -klotho-FGFR1c heterodimers. With FGF21 thus tethered via its C-terminal tail to β -klotho monomers and/or β -klotho-FGFR1c heterodimers, all three components are reduced to two dimensions at the membrane and the weak—but demonstrable—affinity of the FGF core of FGF21 for FGFR1c is sufficient to drive the formation of the activated ternary FGF21-FGFR1c- β -klotho complex via a reduced dimensionality effect on the bivalent binding of FGF to two FGFR molecules²⁵. In this model, β -klotho

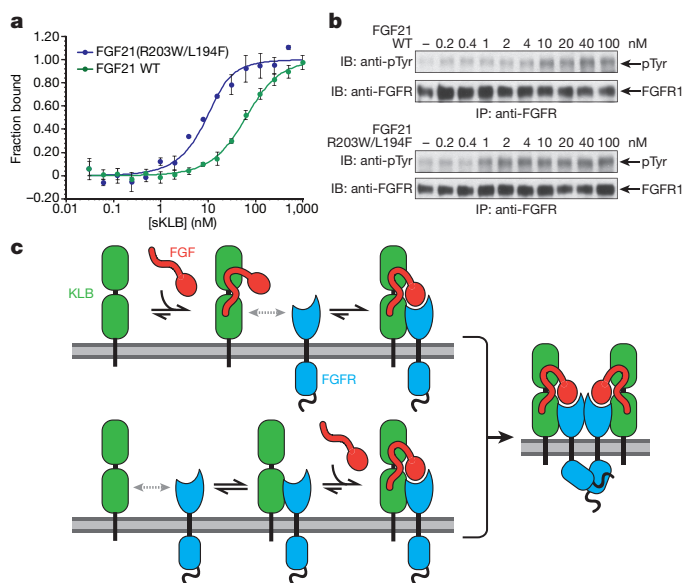


Figure 4 | Structure-based engineering of an FGF21 analogue with enhanced biological activity and the mechanism of endocrine FGF signalling. **a, b**, Enhanced binding affinity (**a**) and bioactivity (**b**) of an FGF21 mutant. Microscale thermophoresis (MST) binding measurements of FGF21 with Leu194Phe and Arg203Trp mutations in the C-terminal tail reveal an approximately tenfold increase in binding affinity to sKLB, with a K_d value of 3.4 ± 1.3 nM, and an approximately tenfold enhanced potency for stimulation of FGFR1c tyrosine phosphorylation. The dots and error bars in **a** denote mean and s.d. of ΔF_{norm} ($n = 3$ independent samples). Individual experimental data are plotted in Extended Data Fig. 9. **c**, A zip-code-like mechanism for β -klotho-dependent FGF21 stimulation of FGFR1c. In the cell membrane of unstimulated cells, β -klotho and FGFR1c monomers are in equilibrium with β -klotho-FGFR heterodimers. Owing to reduced dimensionality, the binding of FGF21 to β -klotho via the FGF21 C-terminal tail, and the bivalent binding of the FGF core of FGF21 to two FGFR1c molecules, will shift the equilibrium towards the formation of an FGF21-FGFR1c- β -klotho ternary complex, and result in the stimulation of tyrosine kinase activity and cell signalling via FGFR1c. In addition, β -klotho functions as a primary high-affinity receptor for FGF21, and FGFR1c functions as a catalytic subunit that mediates receptor dimerization and intracellular signalling.

functions as a primary high-affinity receptor for FGF21, whereas FGFR1c functions as a catalytic subunit that mediates receptor dimerization and intracellular signalling.

The crystal structure of sKLB bound to FGF21_{CT} also provides clear evidence for how the two glycoside hydrolase-like domains of β -klotho have been ‘repurposed’ in evolution to recognize FGF21 specifically. By comparing the structures of substrate-bound β -glucosidases to the second glycoside hydrolase-like domain of FGF21_{CT}-bound β -klotho, we reveal how the active site of an enzyme specialized for cutting sugars has evolved to become a specific and high-affinity cell-surface receptor for circulating hormones that regulate essential metabolic processes including the lowering of blood sugar levels—this may not be a coincidence. The structure of the C terminus of FGF21 appears to mimic that of an oligosaccharide. The similarities between FGF21 and FGF19 indicate that the specificity of these two hormones towards β -klotho, and their modes of action, are very similar (Extended Data Fig. 5). Differences in the cellular responses to these two endocrine FGFs are likely to be determined by the altered binding preferences of the two ligands for the different FGFRs.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 July; accepted 3 November 2017.

Published online 17 January 2017.

1. Eswarakumar, V. P., Lax, I. & Schlessinger, J. Cellular signaling by fibroblast growth factor receptors. *Cytokine Growth Factor Rev.* **16**, 139–149 (2005).
2. Belov, A. A. & Mohammadi, M. Molecular mechanisms of fibroblast growth factor signaling in physiology and pathology. *Cold Spring Harb. Perspect. Biol.* **5**, a015958 (2013).
3. Ogawa, Y. *et al.* β Klotho is required for metabolic activity of fibroblast growth factor 21. *Proc. Natl Acad. Sci. USA* **104**, 7432–7437 (2007).
4. Urakawa, I. *et al.* Klotho converts canonical FGF receptor into a specific receptor for FGF23. *Nature* **444**, 770–774 (2006).
5. Owen, B. M., Mangelsdorf, D. J. & Kliewer, S. A. Tissue-specific actions of the metabolic hormones FGF15/19 and FGF21. *Trends Endocrinol. Metab.* **26**, 22–29 (2015).
6. Koshland, D. E. Jr. Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. Camb. Philos. Soc.* **28**, 416–436 (1953).
7. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
8. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
9. Hayward, S. & Lee, R. A. Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J. Mol. Graph. Model.* **21**, 181–183 (2002).
10. Yie, J. *et al.* FGF21 N- and C-termini play different roles in receptor interaction and activation. *FEBS Lett.* **583**, 19–24 (2009).
11. Micanovic, R. *et al.* Different roles of N- and C- termini in the functional activity of FGF21. *J. Cell. Physiol.* **219**, 227–234 (2009).
12. Hecht, R. *et al.* Rationale-based engineering of a potent long-acting FGF21 analog for the treatment of type 2 diabetes. *PLoS ONE* **7**, e49345 (2012).
13. Zhen, E. Y., Jin, Z., Ackermann, B. L., Thomas, M. K. & Gutierrez, J. A. Circulating FGF21 proteolytic processing mediated by fibroblast activation protein. *Biochem. J.* **473**, 605–614 (2016).
14. Dunshee, D. R. *et al.* Fibroblast activation protein cleaves and inactivates fibroblast growth factor 21. *J. Biol. Chem.* **291**, 5986–5996 (2016).
15. Coppage, A. L. *et al.* Human FGF-21 is a substrate of fibroblast activation protein. *PLoS ONE* **11**, e0151269 (2016).
16. Chuenchor, W. *et al.* The structural basis of oligosaccharide binding by rice BGlu1 β -glucosidase. *J. Struct. Biol.* **173**, 169–179 (2011).
17. Isorna, P. *et al.* Crystal structures of *Paenibacillus polymyxa* β -glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family I glycosidases. *J. Mol. Biol.* **371**, 1204–1218 (2007).
18. Degirolamo, C., Sabbà, C. & Moschetta, A. Therapeutic potential of the endocrine fibroblast growth factors FGF19, FGF21 and FGF23. *Nat. Rev. Drug Discov.* **15**, 51–69 (2016).
19. Kharitonov, A. *et al.* Rational design of a fibroblast growth factor 21-based clinical candidate, LY2405319. *PLoS ONE* **8**, e58575 (2013).
20. Huang, Z. *et al.* A better anti-diabetic recombinant human fibroblast growth factor 21 (rhFGF21) modified with polyethylene glycol. *PLoS ONE* **6**, e20669 (2011).
21. Huang, J. *et al.* Development of a novel long-acting antidiabetic FGF21 mimetic by targeted conjugation to a scaffold antibody. *J. Pharmacol. Exp. Ther.* **346**, 270–280 (2013).
22. Foltz, I. N. *et al.* Treating diabetes and obesity with an FGF21-mimetic antibody activating the β Klotho/FGFR1c receptor complex. *Sci. Transl. Med.* **4**, 162ra153 (2012).
23. Luo, J. *et al.* A nontumorigenic variant of FGF19 treats cholestatic liver diseases. *Sci. Transl. Med.* **6**, 247ra100 (2014).
24. Kolumam, G. *et al.* Sustained brown fat stimulation and insulin sensitization by a humanized bispecific antibody agonist for fibroblast growth factor receptor 1/ β Klotho complex. *EBioMedicine* **2**, 730–743 (2015).
25. Schlessinger, J. *et al.* Crystal structure of a ternary FGF–FGFR–heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol. Cell* **6**, 743–750 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements The NSLS-SSRL is supported by P41GM111244, P41GM103393, DE-SC0012704 and by DE-AC02-76SF00515. We thank NE-CAT (P41 GM103403) and APS (DE-AC02-06CH11357). This research was also supported by NIH grant 1S100D018007 and NIH Award S10RR026992-0110. J.St. thanks INSTRUCT (ESFRI, FWO) for financial support and I. Aboutaleb for technical assistance.

Author Contributions S.L. designed, performed experiments and determined the crystal structures. J.C., J.M. and F.T. provided technical support. E.P. and J.St. generated nanobodies. L.P.S. and I.L. designed and analysed cell-based experiments. S.L., M.A.L. and J.Sc. designed experiments, analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.Sc. (joseph.schlessinger@yale.edu).

Reviewer Information Nature thanks N. Jura, K. White, H. E. Xu and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Plasmid construction. cDNAs that encode for either amino acids 30–983 (sKLB) or 30–522 (KLB_{D1}) of human β -klotho (KLB) were amplified together with the tobacco etch virus (TEV) protease cleavage site and linker of four Gly residues. The resulting sequence was subcloned into a modified pCEP4 vector (Thermo Fisher Scientific) that contains the sequence for the Fc region of human IgG1. The expression vector for C-terminal HA-tagged KLB was generated by subcloning the gene of full-length KLB together with the haemagglutinin (HA)-tag sequence into a pBABE vector. All plasmids of KLB mutants were generated by following standard site-directed mutagenesis protocol using a plasmid containing wild-type C-terminal HA-tagged KLB.

Expression and purification of sKLB and KLB_{D1}. HEK293 EBNA cells were cultured in a humidified incubator with 5% CO₂ at 37°C in DMEM (Thermo Fisher Scientific) containing 10% fetal bovine serum (FBS), 100 U ml⁻¹ penicillin–streptomycin, and 250 µg ml⁻¹ G418. The plasmids were transfected into HEK293 EBNA cells with the Lipofectamine 2000 (Thermo Fisher Scientific) and selected by treatment with 200 µg ml⁻¹ of hygromycin B (Thermo Fisher Scientific) for 2–3 weeks. Cells stably expressing sKLB–Fc or KLB_{D1}–Fc were expanded in Hyperflasks (Corning), and the medium was changed to DMEM with 5% FBS when cell confluency had reached about 70%. After 7 days, the medium was collected after centrifugation at 5,000g and filtration through a 0.2-µm membrane. Swainsonine (15 µM; Cayman Chemical) was added to the medium of cultured cells when preparing proteins for crystallization.

Medium collected from the cells expressing sKLB–Fc or KLB_{D1}–Fc was incubated with recombinant protein A sepharose 4B (Thermo Fisher Scientific) overnight at 4°C. The resin was washed with 50 column volumes of PBS and the protein was eluted from the resin using 0.1 M glycine–HCl, pH 3.5, and immediately neutralized with 0.1 M Tris, pH 7.4. The eluted protein was incubated with recombinant TEV protease for 2 h at room temperature to cleave the C-terminal Fc tag, followed by incubation with recombinant protein A sepharose 4B for 30 min at 4°C to remove Fc tag and undigested protein. The protein was then subjected to cation exchange chromatography (Mono S 5/50 GL, GE Healthcare) using 20 mM sodium phosphate buffer at pH 7.0 (for sKLB) or at pH 6.5 (for KLB_{D1}), and purified using a linear salt gradient. The elution fractions containing sKLB or KLB_{D1} were pooled, concentrated and subjected to a Superdex 200 Increase 10/300 GL (GE Healthcare) size-exclusion chromatography column pre-equilibrated with 20 mM HEPES, 150 mM NaCl, pH 7.0. The eluted fractions containing sKLB or KLB_{D1} were pooled, concentrated, flash-frozen and stored at –80°C until further use. For the crystallization of sKLB, two potential N-glycosylation sites, Asn308 and Asn611, were mutated to glutamine. The mutations were introduced to the sKLB–Fc plasmid by standard QuikChange site-directed mutagenesis. The expression and purification of mutant sKLB was identical to that used for wild-type sKLB. The typical yield of sKLB after complete purification was 1–2 mg per litre of medium from the cells that stably expressed sKLB.

Expression and purification of recombinant FGF21, GST–FGF21_{CT} and FGF1c_{D2D3}. The DNA sequence that encodes for human FGF21 amino acids 29–209 with three mutations—Leu126Arg, Pro199Gly and Asn208Glu—was codon-optimized for *Escherichia coli* expression and synthesized (Blue Heron Biotech). After cloning into a pET28a vector (Novagen), the plasmid was transformed into BL21-Gold (DE3) competent cells. Transformants were grown in LB medium containing 50 µg ml⁻¹ kanamycin, shaken at 240 r.p.m. at 37°C. When the A_{600 nm} of the samples reached 0.6, the bacteria were induced with 1 mM IPTG for 4 h at 37°C. The bacterial cell pellet, collected by centrifugation at 5,000g at 4°C, was lysed in 20 mM sodium phosphate buffer, 500 mM NaCl, 5% glycerol, pH 7.8, using EmulsiFlex-C3 homogenizer (Avestin), followed by centrifugation at 20,000g for 30 min at 4°C. The supernatant containing N-terminal His₆-tagged FGF21 was supplemented with 10 mM imidazole and incubated with Ni-NTA agarose (Qiagen) for 1 h at 4°C. The resin was washed with a 20 column volume of lysis buffer containing 10 mM imidazole, and the protein was eluted from the resin with lysis buffer containing 300 mM imidazole. The protein solution was injected into a HiLoad 26/600 Superdex 200 (GE Healthcare) size-exclusion chromatography column equilibrated with 20 mM HEPES, 900 mM NaCl at pH 7.5. The eluted fractions containing FGF21 were pooled, concentrated to about 1.5 mg ml⁻¹, flash-frozen and stored at –80°C. To generate glutathione S-transferase (GST)-tagged FGF21_{CT}, a DNA sequence encoding amino acids 169–209 of FGF21 was cloned into pGEX-4T-1 vector (GE Healthcare), and the plasmid was transformed into BL21-Gold (DE3) competent cells (Agilent). Transformants were grown in LB medium containing 100 µg ml⁻¹ ampicillin at 37°C until A_{600 nm} reached 0.6, and induced with 1 mM IPTG for 4 h at 37°C. Bacteria cells were collected, lysed in PBS using EmulsiFlex-C3 homogenizer (Avestin), and centrifuged at 20,000g for 30 min at 4°C. The supernatant containing GST–FGF21_{CT} was incubated with glutathione sepharose 4B (GE Healthcare) pre-equilibrated with PBS, for 1 h at 4°C. The beads were washed with 50 column volumes of PBS and the protein was

eluted with 20 mM HEPES, 150 mM NaCl, 10 mM reduced glutathione, pH 7.3. The protein solution containing GST–FGF21_{CT} was then dialysed against 20 mM HEPES, 150 mM NaCl before flash-freezing and storage at –80°C. A peptide corresponding to the C-terminal region of FGF21 containing amino acids 174–209 with two substitutions, Pro199Gly and Ala208Glu, was synthesized and purified by the Tufts University Core Facility. The ligand-binding region of FGF1c was expressed in *E. coli* as an insoluble fraction. The protein was refolded and purified as previously described²⁵.

Expression and purification of Nb914. The plasmid containing C-terminal His₆-tagged Nb914 was transformed into *E. coli* strain WK6, and grown in TB medium containing 0.1% glucose, 2 mM MgCl₂, and 100 µg ml⁻¹ ampicillin at 37°C until the A_{600 nm} of the sample was 1.2, and then induced with 1 mM IPTG for 4 h. Cells were collected and the periplasmic fraction was extracted using the modified osmotic shock protocol²⁶. The periplasmic extract containing Nb914 was supplemented with 10 mM imidazole and incubated with Ni-NTA agarose (Qiagen) for 1 h at 4°C. The beads were washed with 50 column volumes of PBS containing 10 mM imidazole, and Nb914 was eluted from the resin with PBS containing 300 mM imidazole. The eluted fraction containing Nb914 were concentrated and injected into a HiLoad 26/600 Superdex 200 (GE Healthcare) size-exclusion chromatography column pre-equilibrated with PBS at pH 7.0. Purified Nb914 at a concentration of 10 mg ml⁻¹ was flash-frozen and stored at –80°C.

Crystallization, X-ray diffraction data collection and structure determination. Purified sKLB or KLB_{D1} was mixed with Nb914, concentrated and injected into a Superdex 200 Increase 10/300 GL (GE Healthcare) size-exclusion chromatography column pre-equilibrated with 20 mM HEPES, 150 mM NaCl, pH 7.0. Eluted fractions containing the complex were pooled, concentrated to 7 mg ml⁻¹ and screened for crystallization using Mosquito Crystal liquid handler (TTP Labtech). Ninety-six-well plates were incubated and imaged at 20°C using Rock Imager 1000 (Formulatrix). sKLB in complex with Nb914 produced rod-shaped crystals when mixed with an equal volume of well solution containing 14% PEG4000, 0.1 M MES, pH 6.0 and equilibrated for 10–15 days using the hanging-drop vapour diffusion method. The crystals were cryopreserved by gradually transferring crystals to the mother liquor supplemented with 30% glucose before being flash-frozen in liquid nitrogen. KLB_{D1} in complex with Nb914 gave plate-like crystals when mixed with an equal volume of well solution containing 30% PEG1000, 0.1 M HEPES pH 7.5 and equilibrated for 4–6 days using the hanging-drop vapour diffusion method; these crystals were directly flash-frozen in liquid nitrogen. For sKLB in complex with Nb914 and FGF21_{CT}, FGF21_{CT} was dissolved in 14% PEG4000, 0.1 M MES, pH 6.0 and added to the drop containing crystals of sKLB. The addition of FGF21_{CT} immediately caused deformation in most of the crystals. Crystals that stayed intact were gradually transferred into the artificial mother liquor, supplemented with 30% glucose and 50 µM FGF21_{CT}, before being flash-frozen in liquid nitrogen. X-ray diffraction data were collected at the beamline BL-14 at the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory (for KLB_{D1} and sKLB) and 24-ID-E at the Advanced Photon Source, Argonne (for sKLB in complex with FGF21_{CT}). The diffraction datasets were processed using HKL2000²⁷ and XDS²⁸. Initial phases for the dataset for KLB_{D1} in complex with Nb914 were calculated by molecular replacement with PHASER²⁹ using the coordinates of the cytosolic β -glucosidase (PDB code: 2ZOX) and the coordinates of a nanobody that exhibits the highest sequence identities with Nb914 (PDB code: 5IMK, chain B) as the search models. Refinement was iteratively performed using PHENIX³⁰ followed by manual model building using Coot³¹. The final coordinates of KLB_{D1} in complex with Nb914 were then used as a search model for the dataset of sKLB in complex with Nb914, together with the coordinates for KLB_{D1}, as a search model for D2 of sKLB. Then, the model was iteratively built and refined for sKLB. For the dataset for sKLB in complex with Nb914 and FGF21_{CT}, initial phase information was obtained by molecular replacement using the final coordinates of sKLB in complex with Nb914, which were divided into two models each containing the coordinates for D1 with Nb914 and D2, and searched independently. Iterative cycles of refinement and rebuilding of the sKLB model improved the phase, and produced significant electron densities for FGF21_{CT}. Subsequently, the model for FGF21_{CT} was manually built on the basis of the $|F_o| - |F_c|$ map, followed by the final refinement cycles. The data collection and refinement statistics are summarized in Supplementary Table 1. All the figures containing the structures were generated using the PyMOL Molecular Graphics System, version 1.8 (Schrödinger).

MST measurements. All MST measurements were performed using the Monolith NT.115Pico instrument (NanoTemper) with Monolith NT.115 MST Premium Coated Capillaries. Purified FGF21 was fluorescently labelled using the Monolith Protein Labelling Kit RED-NHS (NanoTemper) according to the manufacturer's instructions. Samples for binding-affinity measurements of FGF21 to sKLB were prepared by mixing 35 nM of fluorescently labelled FGF21 (fl-FGF21) with a series of concentrations (0.03–1,000 nM) of purified sKLB in 20 mM HEPES, 150 mM NaCl, pH 7.0, 0.05% Tween-20, 1 mg ml⁻¹ BSA. The thermophoretic movements

of fl-FGF21 in each sample were monitored (LED 20%, IR laser 20%) and the normalized fluorescence intensities (F_{norm}), defined as $F_{\text{hot}}/F_{\text{cold}}$ (where F_{cold} and F_{hot} refer to the fluorescence intensities averaged over a 1-s period before the IR laser was turned on and 29 s after IR laser was turned on, respectively) for each sample were plotted against the concentrations of sKLB. For the competition assays, the thermophoresis of fl-FGF21 was measured for samples in which the concentration of fl-FGF21 and sKLB mixture was kept constant, with concentrations of GST-FGF21_{CT} varying from 2.1 nM to 35,000 nM. All the data were analysed with the MO.Affinity Analysis software (NanoTemper) provided by the manufacturer. **Surface plasmon resonance measurements.** All surface plasmon resonance experiments were performed using a BIAcore T100 instrument (GE Healthcare) at 25 °C (Keck Foundation Biotechnology Resource Laboratory) using HPBS + buffer (GE Healthcare). Anti-GST antibody (GE Healthcare) was immobilized on a CM5 sensor chip using the instructions provided, followed by capturing 50 response units (RU) of GST-FGF21_{CT}. Using the single-cycle kinetics method, a series concentration of sKLB ranging from 25.6 nM to 1,000 nM was subsequently injected onto the surfaces with 360 s of association period, followed by the dissociation period of 1,200 s. The binding kinetics were evaluated using BIAevaluation software (GE Healthcare).

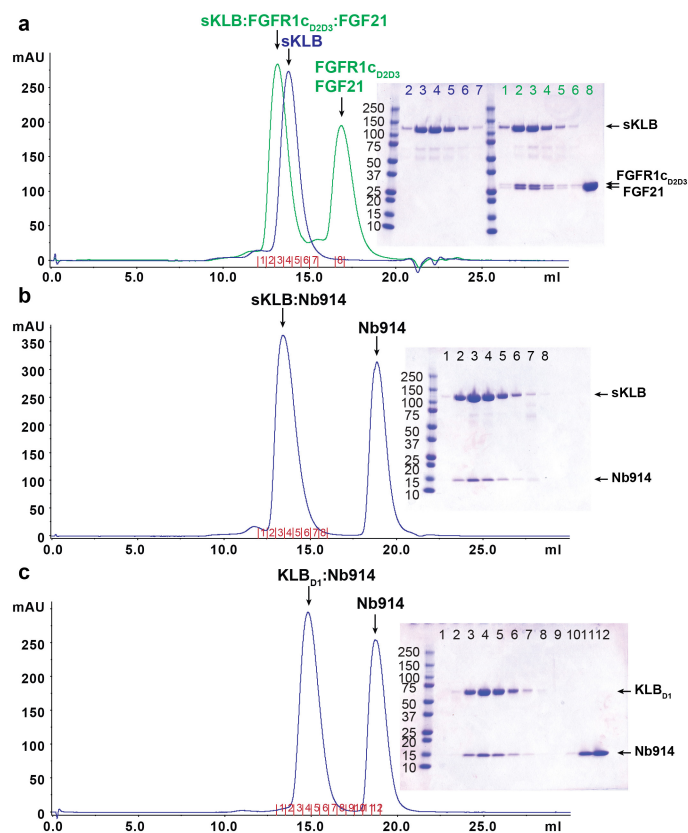
Cell-based activity assays. L6 cells that stably co-expressed wild-type FGFR1c, together with either wild-type β -klotho or a variety of β -klotho mutants, were grown in DMEM supplemented with 10% FBS, 100 U ml⁻¹, penicillin-streptomycin, 0.1 mg ml⁻¹ hygromycin and 1 μ g ml⁻¹ puromycin. Cells were starved overnight in DMEM with 0.5% FBS and stimulated for 10 min at 37 °C

with either FGF1 or FGF21 at concentrations of 5 nM and 25 nM, respectively. Cells were then lysed and subjected to immunoprecipitation with anti-FGFR1 antibody, followed by SDS-PAGE. The samples were then subjected to immunoblotting with anti-phosphotyrosine (pTyr), anti- β -klotho or anti-FGFR1 antibodies.

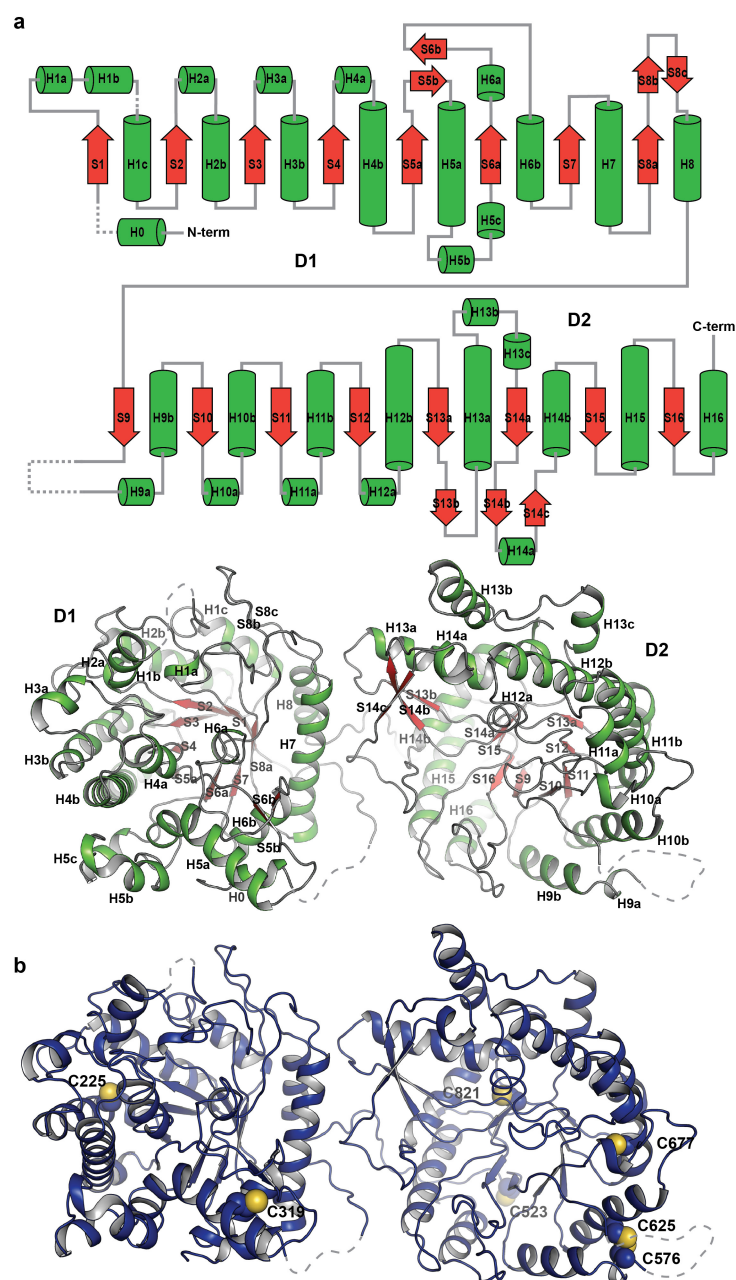
Statistics and reproducibility. No statistical methods were used to predetermine sample size. All of the immunoblots and binding affinity measurements presented in this work were repeated at least three times with similar results.

Data availability. Coordinates and structure factors for the complexes have been deposited in the Protein Data Bank (PDB) under accessions 5VAK (KLB_{D1}-Nb914), 5VAN (sKLB-Nb914) and 5VAQ (sKLB-FGF21_{CT}-Nb914). All other data are available from the corresponding author upon reasonable request.

26. Pardon, E. *et al.* A general protocol for the generation of nanobodies for structural biology. *Nat. Protoc.* **9**, 674–693 (2014).
27. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
28. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
29. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
30. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
31. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
32. Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786 (2011).

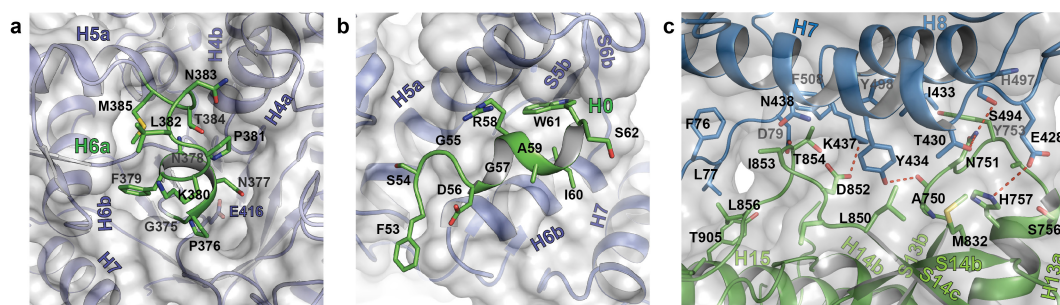


Extended Data Figure 1 | Expression, purification and crystallization of β -klotho extracellular domain. a–c, Size-exclusion chromatography profiles and corresponding Coomassie-stained SDS–PAGE gels of the sKLB–FGFR1c_{D2D3}–FGF21 ternary complex (green) or sKLB alone (blue) (a), sKLB in complex with Nb914 (b) and KLB_{D1} in complex with Nb914 (c). The chromatograms and the SDS–PAGE gels shown are representatives of at least three independent preparations with similar results. A secreted protein composed of the extracellular domain of KLB fused to the Fc region of human IgG1 was produced by HEK293 EBNA cells. Following purification using a protein A agarose resin, the KLB–Fc fusion protein was subjected to proteolytic cleavage. sKLB was further purified using ion exchange and size-exclusion chromatography. Multiple crystallization trials with the ternary complex formed by sKLB, FGF21 and FGFR1c_{D2D3} (a, green) failed to yield diffraction-quality crystals. However, a preparation of sKLB bound to a nanobody Nb914 (b) yielded crystals that diffracted X-rays to a resolution of 6–8 Å, and these were further improved by mutating two of the eleven potential *N*-glycosylation sites in sKLB (Asn308 and Asn611) to glutamine residues. The resulting crystals of an sKLB–Nb914 complex diffracted to a resolution of 2.2 Å. We also crystallized KLB_{D1} in complex with Nb914 (c), and collected data to a resolution of 1.7 Å. The structure of KLB_{D1} was first solved by molecular replacement using the coordinates of a structure of human cytosolic β -glucosidase (PDB code: 2ZOX) and the coordinates of a nanobody structure (PDB code: 5IMK, chain B) as search models. The structure of sKLB was subsequently determined by molecular replacement using the KLB_{D1} coordinates as a search model.



Extended Data Figure 2 | Domain diagram of sKLB structure and the location of cysteine residues. **a**, Secondary structure elements (H for helix (green) and S for sheet (red)) are designated by numbers on the basis of the principal elements for the $(\beta/\alpha)_8$ fold. Dashed lines depict disordered loops that are not modelled in the structure. **b**, Seven of the ten cysteine residues in the extracellular region were successfully modelled in the sKLB structure. With the exception of the disulfide bond between Cys576 and Cys625, the structure shows that these cysteine residues are reduced and do not form disulfide bridges. Moreover, determination of the distances between each pair of cysteines indicates that most are too

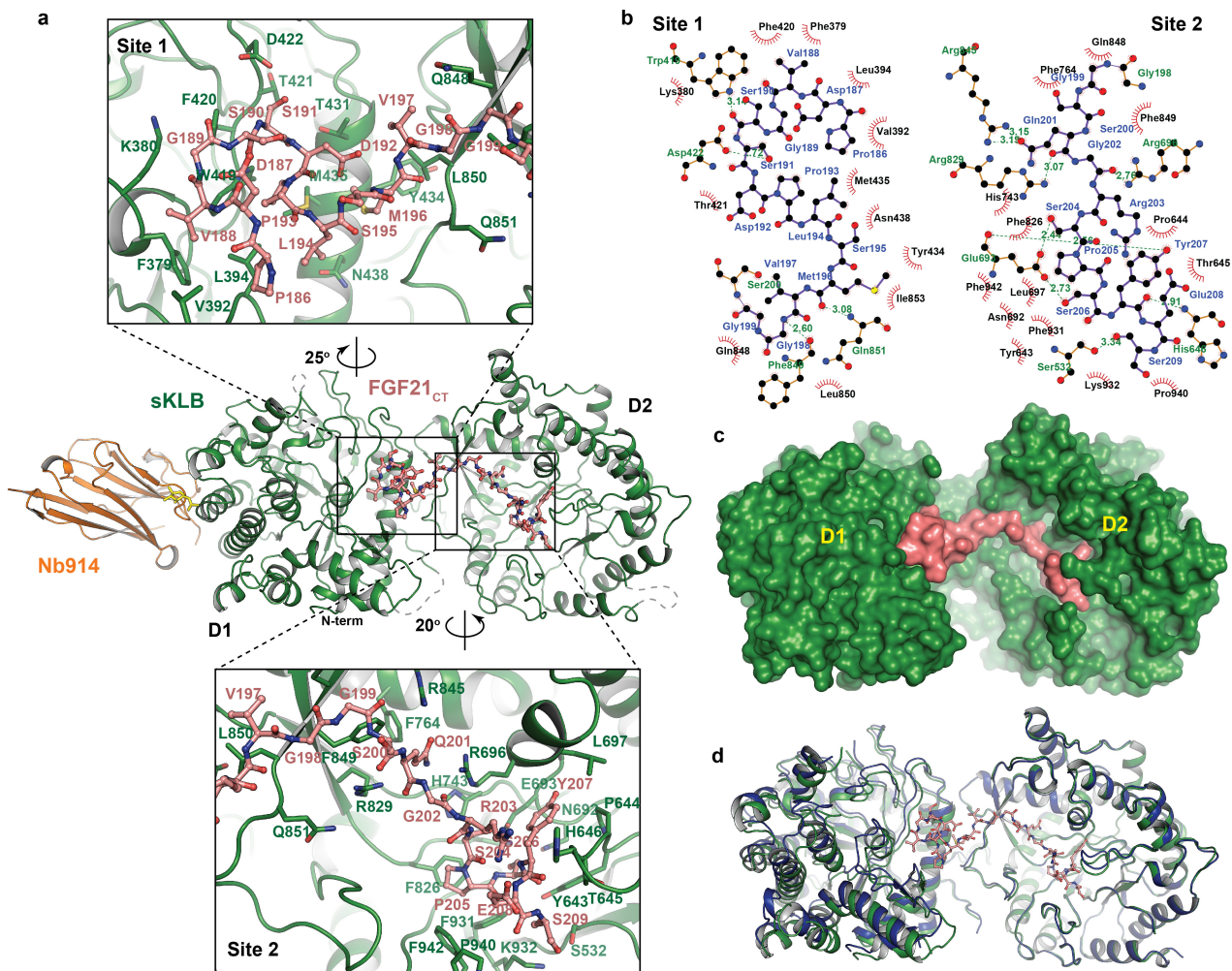
far apart to form intramolecular disulfide bonds. However, we cannot rule out the possibility that Cys976 located in the C-terminal region of sKLB, which could not be modelled owing to weak electron density, may form a disulfide bond with the nearby Cys523. There is no evidence for the formation of intermolecular disulfide bonds between β -klotho and the closely associated FGFR, FGF19 or FGF21 proteins, whose cysteines all form well-characterized intramolecular disulfide bonds. The functional consequences of the presence of reduced cysteines in β -klotho are currently unknown.



Extended Data Figure 3 | Unique structural features of sKLB.

a, Interaction of H6a (green) with the pseudo-substrate binding pocket in D1 of sKLB. Glu416, the pseudo-catalytic glutamic acid residue in D1, is located on the bottom of the pocket and is also highlighted. **b**, Interaction

of H0 (green) with the nearby structural elements in D1 of sKLB. **c**, Interface between D1 (blue) and D2 (green) of sKLB, highlighting amino acids and structural elements as well as polar interactions (red dotted lines) between the domains.



Extended Data Figure 4 | Details of interactions between sKLB and FGF21_{CT}, and conformational changes upon ligand binding.

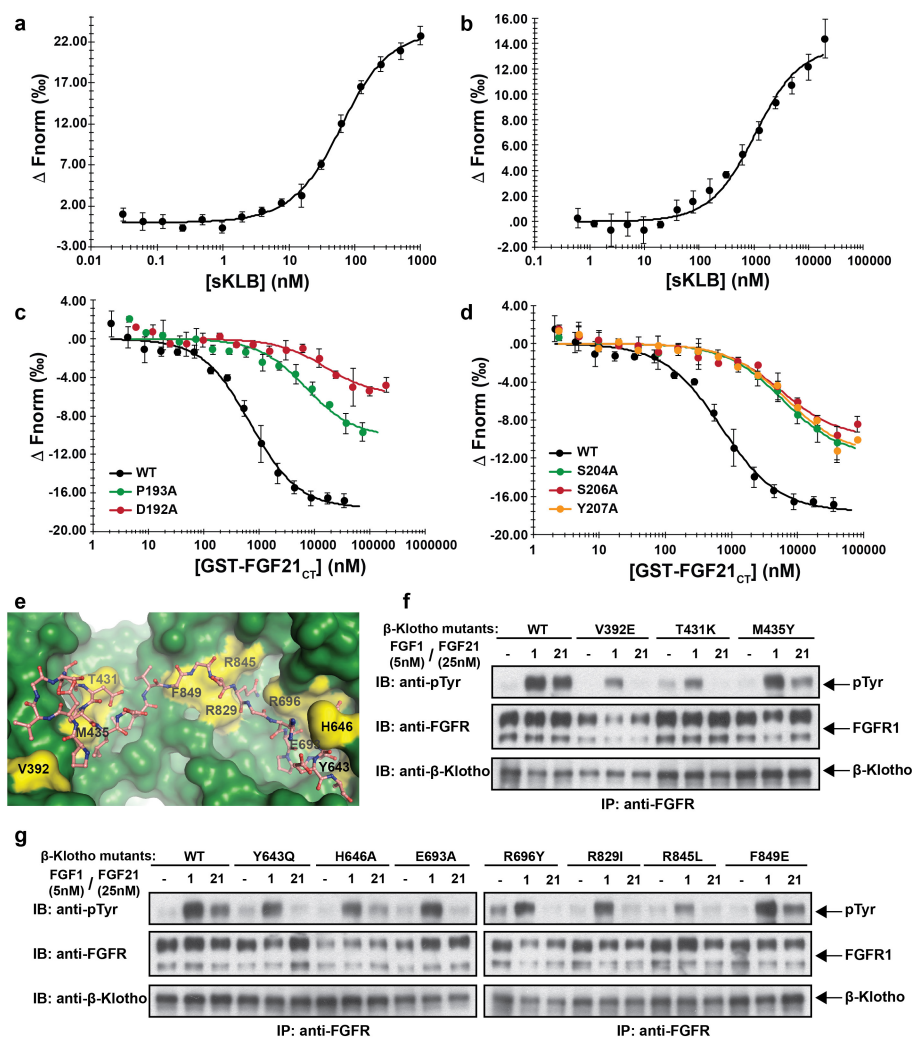
a, Interactions between amino acid residues in sKLB (green) and FGF21_{CT} (salmon) in the areas of sites 1 and 2 are indicated. **b**, Diagram of amino-acid-specific interactions between sKLB and FGF21_{CT} within sites 1 and 2.

The figure was generated using Ligplot+³². **c**, Structure of sKLB (green) in complex with FGF21_{CT} (salmon) shown as a surface representation. **d**, Structure of ligand-free sKLB (blue) is overlaid onto the structure of sKLB (green) bound to FGF21_{CT} (salmon, ball-and-stick).

FGF19 (189-216)

190 200
 PQPPDVGSSDPLSMVG-PSQGRSPSYAS
 190 200 210
 SSPLETDSMDPFGLVGTGLEAVRSPSFEK

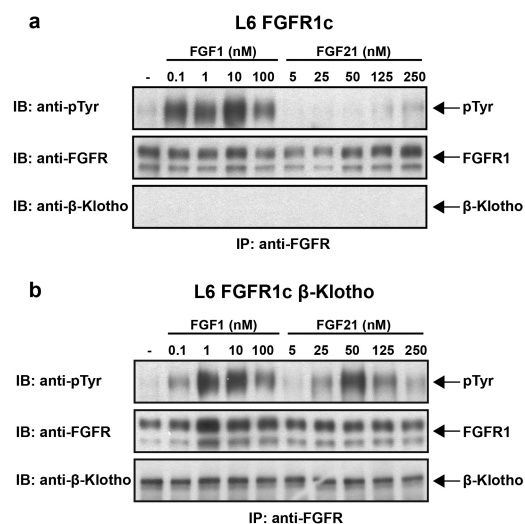
Extended Data Figure 5 | Amino acid sequence alignments of C-terminal regions of human FGF19 and FGF21. Residues Asp-Pro, which are critical in maintaining multi-turn elements, are highlighted in blue, and the sugar-mimicking motif Ser-Pro-Ser is highlighted in yellow. The sequence alignment reveals close sequence similarity between the C-terminal tails of FGF21 and FGF19 that is consistent with the similar binding characteristics of FGF21 and FGF19 and their isolated C-terminal regions to β -klotho. The sugar-mimicking motif in FGF21, Ser205-Pro206-Ser207, is conserved in FGF19 (Ser211-Pro212-Ser213). The sequence Asp192-Pro193, in the region of FGF21_{CT} that binds to site 1 of β -klotho by stabilizing intramolecular hydrogen bonds that maintain a turn in the bound configuration of FGF21_{CT}, is also highlighted. This sequence is conserved in FGF19 (Asp198-Pro199), which suggests that intramolecular interactions similar to those responsible for mediating consecutive turns in FGF19_{CT} may also bind to site-1 of β -klotho. Because many of the intramolecular interactions within FGF21_{CT} bound to β -klotho take place between main-chain atoms (as observed in typical β -turn structures), the presence of only a few key amino acid sequences such as Asp198-Pro199 may be sufficient to generate multi-turn elements in FGF19_{CT} that are similar to those observed in the crystal structure of FGF21_{CT} bound to β -klotho.



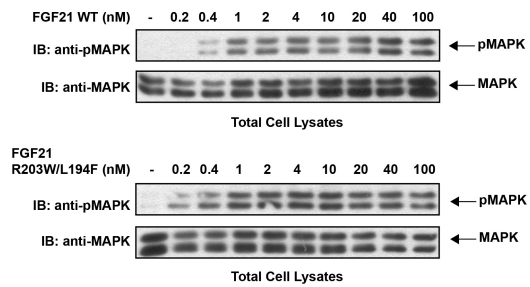
Extended Data Figure 6 | Validation of FGF21-binding interface to β-klotho by ligand-binding and cell-stimulation experiments.

a, b, MST-based binding affinity measurements of (a) FGF21 to sKLB and FGFR1_{C_D2D₃} to sKLB (b) that yielded $K_d = 43.5 \pm 5.0$ nM and $K_d = 940 \pm 176$ nM, respectively. **c, d**, MST-based competition assay with GST-FGF21_{CT} that contained mutations in regions that interact with site 1 (c) or site 2 (d). Half-maximal inhibitory concentration (IC_{50}) values for wild type, 704 ± 96 nM; D192A, $15,900 \pm 6,210$ nM; P193A, $7,160 \pm 2,350$ nM; S204A, $5,990 \pm 1,040$ nM; S206A, $5,560 \pm 1,590$ nM; and Y207A, $6,630 \pm 1,570$ nM. The dots and error bars in panels **a–d**

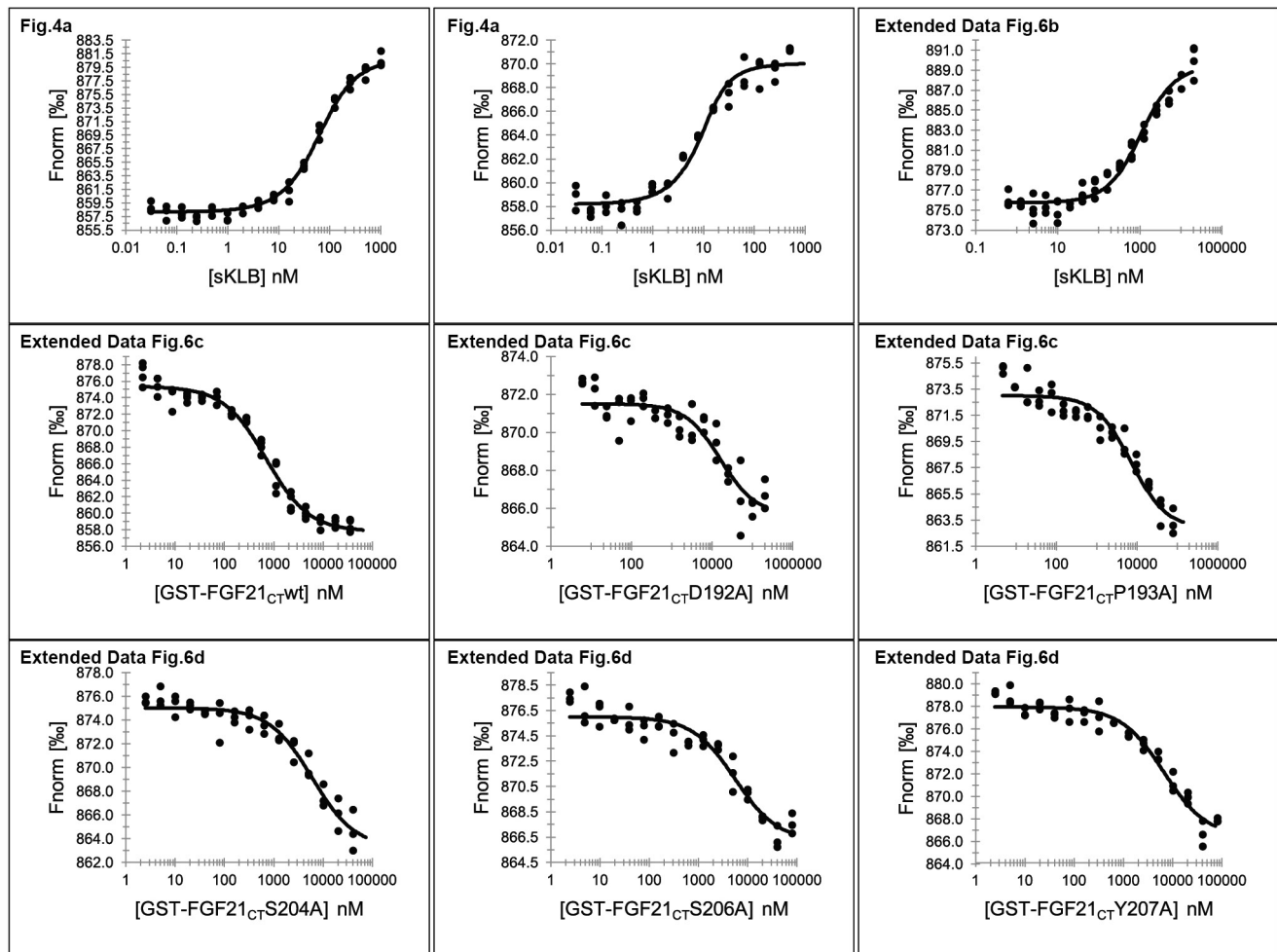
denote mean and s.d. of ΔF_{norm} ($n = 3$ independent samples). Individual experimental data are plotted in Extended Data Fig. 9. **e**, Location of mutated amino acid residues (yellow) in sKLB (green) occupied by FGF21 (salmon) that were analysed in **f** and **g**. **f, g**, Stably transfected L6 cells co-expressing FGFR1c together with wild-type or β-klotho mutants were stimulated with either FGF21 or FGF1 (control) and analysed for FGFR1c activation by monitoring tyrosine phosphorylation of FGFR1c. Lysates of ligand-stimulated or unstimulated cells were subjected to immunoprecipitation with anti-FGFR1 antibodies, followed by immunoblotting with either anti-pTyr or anti-FGFR1 antibodies.



Extended Data Figure 7 | β -Klotho is required for FGFR1c-mediated signalling induced by FGF21. **a, b,** L6 cells that expressed either FGFR1c alone (**a**) or FGFR1c together with β -klotho (**b**) were stimulated with a range of concentrations of FGF1 or FGF21, and phosphotyrosine (pTyr) levels of FGFR were monitored by immunoprecipitation with anti-FGFR1 antibodies, followed by immunoblotting with anti-pTyr antibodies.



Extended Data Figure 8 | MAP kinase stimulation induced by wild-type or mutant FGF21. L6 cells that co-expressed β -klotho and FGFR1c were stimulated with wild-type FGF21 (top) or FGF21(R203W/L194F) (bottom), and phosphorylation levels of MAP kinase in cell lysates were monitored.



Extended Data Figure 9 | MST data with individual data points. Figures that contain the data are indicated.

Extended Data Table 1 | Crystallographic data collection and refinement statistics

	KLB _{D1} :Nb914	sKLB:Nb914	sKLB:Nb914:FGF21 _{CT}
Data collection			
Space group	C2	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions			
a, b, c (Å)	229.43, 49.35, 54.31	48.68, 144.07, 215.61	48.65, 145.49, 213.83
α, β, γ (°)	90, 100.22, 90	90, 90, 90	90, 90, 90
Resolution (Å)	41.27-1.70 (1.76-1.70)	47.49-2.20 (2.28-2.20)	60.14-2.61 (2.70-2.61)
R _{merge}	0.0546 (0.441)	0.109 (0.880)	0.0905 (1.289)
CC _{1/2} (%)	99.9 (86.4)	99.7 (76.3)	99.7 (43.1)
<I/σ>	20.36 (2.87)	20.13 (2.71)	13.12 (1.03)
Completeness (%)	98 (87)	100 (97)	98 (96)
Redundancy	4.2 (3.4)	7.1 (6.9)	3.9 (3.3)
Refinement			
Resolution (Å)	41.27-1.70	47.49-2.20	60.14-2.61
No. of reflections used	65178	77784	46521
R _{work} / R _{free} (%)	17.16 / 19.64	18.62 / 21.06	19.11 / 22.89
No. of atoms			
Protein	4453	7731	7874
Ligands	42	125	13
Waters	271	112	0
Average B-factors			
Protein	26.11	43.00	63.42
Ligands	45.67	65.37	68.90
Waters	30.97	39.59	n/a
R.m.s. deviations			
Bond lengths (Å)	0.006	0.008	0.012
Bond angle (°)	0.82	0.95	1.30

Regulation of embryonic haematopoietic multipotency by EZH1

Linda T. Vo^{1,2,3}, Melissa A. Kinney^{1,2}, Xin Liu⁴, Yuannyu Zhang^{4,5}, Jessica Barragan^{1,2}, Patricia M. Sousa^{1,2}, Deepak K. Jha^{1,2}, Areum Han^{1,2}, Marcella Cesana^{1,2}, Zhen Shao⁵, Trista E. North⁶, Stuart H. Orkin^{2,3,7}, Sergei Doulatov⁸, Jian Xu⁴ & George Q. Daley^{1,2,3}

All haematopoietic cell lineages that circulate in the blood of adult mammals derive from multipotent haematopoietic stem cells (HSCs)¹. By contrast, in the blood of mammalian embryos, lineage-restricted progenitors arise first, independently of HSCs, which only emerge later in gestation^{2,3}. As best defined in the mouse, ‘primitive’ progenitors first appear in the yolk sac at 7.5 days post-coitum^{2,3}. Subsequently, erythroid–myeloid progenitors that express fetal haemoglobin⁴, as well as fetal lymphoid progenitors⁵, develop in the yolk sac and the embryo proper, but these cells lack HSC potential. Ultimately, ‘definitive’ HSCs with long-term, multilineage potential and the ability to engraft irradiated adults emerge at 10.5 days post-coitum from arterial endothelium in the aorta-gonad-mesonephros and other haemogenic vasculature³. The molecular mechanisms of this reverse progression of haematopoietic ontogeny remain unexplained. We hypothesized that the definitive haematopoietic program might be actively repressed in early embryogenesis through epigenetic silencing⁶, and that alleviating this repression would elicit multipotency in otherwise lineage-restricted haematopoietic progenitors. Here we show that reduced expression of the Polycomb group protein EZH1 enhances multi-lymphoid output from human pluripotent stem cells. In addition, *Ezh1* deficiency in mouse embryos results in precocious emergence of functional definitive HSCs *in vivo*. Thus, we identify EZH1 as a repressor of haematopoietic multipotency in the early mammalian embryo.

The differentiation of pluripotent stem cells to haematopoietic lineages generates robust erythroid–myeloid lineage-restricted progenitors but not HSCs. This pattern bears marked similarities to early haematopoietic ontogeny. We hypothesized that the same epigenetic factors actively repress multipotency in embryogenesis and differentiation from pluripotent stem cells. To identify these factors, we adopted a loss-of-function screen using lentivirally delivered short hairpin RNAs (shRNAs) that target 20 DNA- and histone-modifying factors (Extended Data Fig. 1a, Supplementary Table 1). Erythroid–myeloid progenitors differentiated from human pluripotent stem cells marked by CD34 and CD45 were expanded with five transcription factors (5F). They retained embryonic features, including lack of lymphoid potential⁷, and this enabled us to screen for reactivation of lymphoid potential as a measure of multipotency. 5F cells were transduced with individual shRNAs and screened for T cell potential on OP9-DL1 stromal cells (Fig. 1a). The knockdown of six factors independently enhanced CD4⁺CD8⁺ T cell potential from 5F cells (Fig. 1b, Extended Data Fig. 1b).

Prospective validation revealed that only *EZH1* knockdown (using shEZH1) elicited robust T (16.3 ± 7.4%; mean ± s.e.m.) and B (22.5 ± 7.3%) cell potential (Fig. 1c–e), compared to shRNAs targeting a control luciferase gene (shLUC) (T cell 0.002 ± 0.002%; B cell

0.022 ± 0.006%) across multiple induced pluripotent stem (iPS) cell lines (Fig. 1f). *EZH1*-deficient cells retained erythroid–myeloid potential as shown by colony-forming assays (Fig. 1g) and flow cytometry (Fig. 1h, i). *EZH1* knockdown also promoted lymphoid potential independently of the five transcription factors, as evidenced by robust T cell differentiation from naive CD34⁺ haemogenic endothelial cells (26.1 ± 16.5% shEZH1 versus 2.3 ± 0.4% shLUC) (Extended Data Fig. 1c). Further characterization was not possible owing to the limited proliferation of pluripotent stem and haemogenic endothelial cells. By contrast, 5F cells expanded exponentially (Extended Data Fig. 1d) and showed increased CD34⁺ progenitors after shEZH1 transduction (78.8 ± 14.2% versus 29.3 ± 10.0%) (Extended Data Fig. 1e). Taken together, these data show that *EZH1* knockdown activates multipotency in lineage-restricted embryonic haematopoietic progenitors.

EZH1 is a component of the Polycomb repressive complex 2 (PRC2), which mediates epigenetic silencing of genes via methylation of lysine residue 27 of histone H3⁸. To dissect the role of PRC2 in repressing haematopoietic multipotency, we assessed T cell differentiation upon depletion of each PRC2 subunit. In addition to *EZH1*, *SUZ12* knockdown also enhanced T cell potential, albeit to a lesser extent. By contrast, knockdown of *EED* or *EZH2* had no effect on T cell potential and dual *EZH1* and *EZH2* knockdown phenocopied that of *EZH2* depletion (Fig. 2a, b). To determine whether the catalytic SET domain was required, we overexpressed full-length mouse *Ezh1* or mutant *Ezh1* lacking the SET domain (mEzh1ΔSET) (Fig. 2c). Overexpression of mouse *Ezh1* completely abrogated T cell potential in shEZH1 cells, whereas the mutant mEzh1ΔSET did not (Fig. 2c, d, Extended Data Fig. 2d–g). Furthermore, overexpression of mouse *Ezh2* failed to suppress T cell potential, despite the remarkable homology of the SET domains (Extended Data Fig. 2e, h, i). These data show that specific inhibition of *EZH1*, rather than antagonism of canonical PRC2, unlocks lymphoid potential and the catalytic SET domain is required for this function.

To understand the molecular changes upon *EZH1* knockdown, we performed RNA sequencing (RNA-seq), assay for transposase-accessible chromatin using sequencing (ATAC-seq) and chromatin immunoprecipitation followed by sequencing (ChIP-seq). Upregulated genes after *EZH1* knockdown were enriched for biological processes such as defence response ($P = 6.8 \times 10^{-9}$), immune response ($P = 1.2 \times 10^{-7}$) and T cell co-stimulation ($P = 0.03$) (Fig. 3a, b). Human haematopoietic gene signatures⁹, such as of HSCs (stem), multi-lymphoid progenitors (MLP) (early lymphoid) and ProB, were highly enriched in shEZH1 cells, consistent with stem and lymphoid potential (Fig. 3c). We also performed RNA-seq and ATAC-seq on emergent haematopoietic stem and progenitor cells (HSPCs) at 10.5 days post-coitum^{10–12} from the yolk sac and aorta-gonad-mesonephros (AGM) of wild-type, *Ezh1*^{+/-} and

¹Stem Cell Program, Boston Children's Hospital, Boston, Massachusetts, USA. ²Division of Hematology/Oncology, Boston Children's Hospital and Dana Farber Cancer Institute, Boston, Massachusetts, USA. ³Harvard Medical School, Boston, Massachusetts, USA. ⁴Children's Medical Center Research Institute, Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, Texas, USA. ⁵Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. ⁶Department of Pathology, Beth Israel-Deaconess Medical Center, Boston, Massachusetts, USA. ⁷Howard Hughes Medical Institute, Boston, Massachusetts, USA. ⁸Division of Hematology, Department of Medicine, University of Washington, Seattle, Washington, USA.

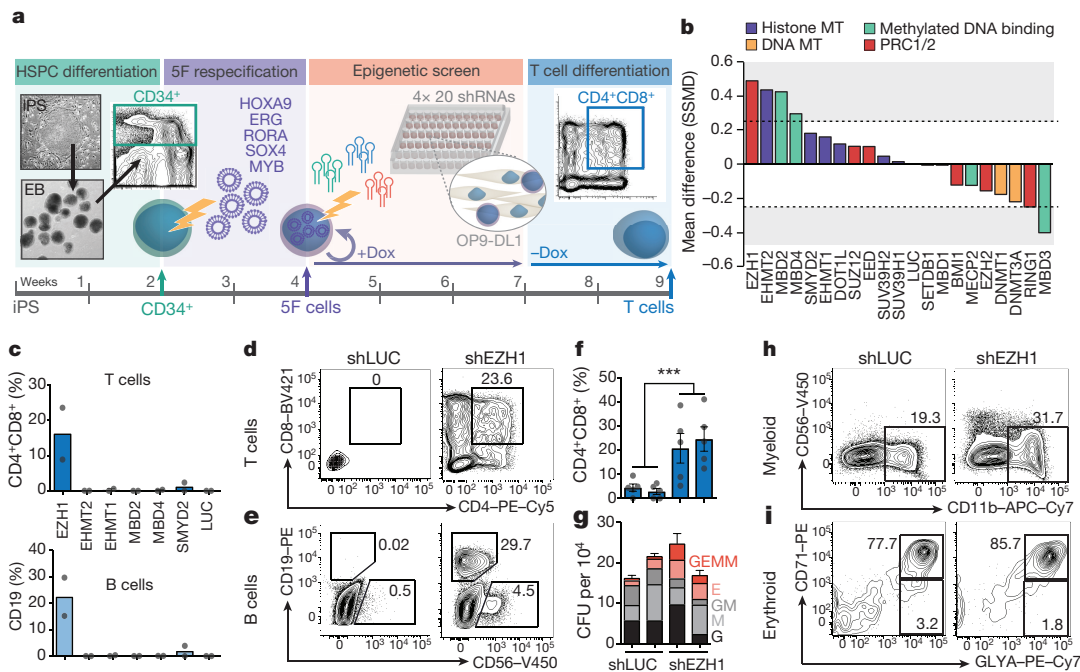


Figure 1 | In vitro screen for epigenetic modifiers that restrict lymphoid potential. **a**, Scheme for human pluripotent stem-cell differentiation into haematopoietic progenitors. CD34⁺ cells were transduced with the five transcription factors (5F) HOXA9, ERG, RORA, SOX4 and MYB. 5F cells were then transduced with individual shRNAs (×4 each) that targeted each epigenetic modifier and seeded onto OP9-DL1 stromal cells to induce T cell differentiation. Dox, doxycycline; EB, embryoid body. **b**, Strictly standardized mean difference (SSMD) of CD4⁺CD8⁺ T cell frequencies across all four shRNAs targeting each epigenetic modifier in 5F cells using two iPS cell lines, CD45-iPS and MSC-iPS1, in two independent experiments. MT, methyltransferase. **c**, Prospective analysis of T and B cell frequencies from 5F cells plus shRNA targeting top candidates ($n = 2$ biological replicates). **d**, Flow analysis of CD4⁺CD8⁺

T cell development of 5F cells with shRNAs targeting luciferase (shLUC) or *EZH1* (shEZH1) after 5 weeks of differentiation on OP9-DL1 stromal cells. **e**, Flow analysis of CD19⁺ B cell potential. **f**, Quantification of T cell potential of 5F plus shEZH1 cells compared to 5F plus shLUC cells pooled across two hairpins and five independent experiments ($n = 10$) using several iPS cell lines (CD34-iPS, CD45-iPS and MSC-iPS1). Individual values obtained for each hairpin are shown in the Source Data. *** $P = 0.001$ by unpaired two-tailed t -test. **g**, Quantification of colony-forming potential in three independent experiments. E, erythroid; GM, granulocyte, monocyte; M, monocyte; G, granulocyte; GEMM, granulocyte, erythroid, monocyte, megakaryocyte. **h, i**, Flow analysis of myeloid (CD11b⁺) (h) and erythroid (CD71⁺GLY⁺) (i) potential. Experiments replicated at least twice. Data are mean \pm s.e.m.

Ezh1^{-/-} mouse embryos (Fig. 4a). Interestingly, in wild-type embryos, the expression of *Ezh1* was lower in the AGM than in the yolk sac, whereas *Ezh2* and *Eed* were higher in the AGM (Fig. 4b). Notably, *Ezh1* deficiency *in vivo* also induced genes enriched for angiogenesis, haematopoietic/lymphoid development and immune system processes (Extended Data Fig. 3a–d).

Regions of increased chromatin accessibility (1,610 ATAC peaks) in shEZH1 cells exhibited concomitantly increased gene expression upon *EZH1* knockdown and were associated with T cell development and lymphocyte activation pathways, as well as HSC, HSC/MLP, B and T cell signatures (Fig. 3d, e, Extended Data Fig. 3e–g). *EZH1* knockdown also increased accessibility to HSC/lymphoid transcription factors, such as *HLF*, *FOXO1* and *ARID5B*^{13–15} (Fig. 3f). Downregulated peaks were enriched for alternative developmental processes and importantly, embryonic haematopoiesis (Fig. 3e, Extended Data Fig. 3e). *In vivo*, upregulated ATAC peaks in *Ezh1*-deficient AGM cells were enriched for immune response, T cell activation, lymphocyte differentiation pathways, as well as HSC and HSC/MLP signatures (Fig. 4a, c, d, Extended Data Fig. 3h, i); furthermore, *Ezh1* deficiency increased accessibility to target genes of master haematopoietic transcription factors, including *Runx1* (Extended Data Fig. 3k, l).

We hypothesized that these molecular changes upon *EZH1* knockdown were mediated by bivalent, or poised, chromatin domains, often implicated in the control of developmentally regulated genes¹⁶. Consistent with previous reports, *EZH1* was broadly associated with repressive (H3K27me3), bivalent (H3K27me3 and H3K4me3) and active (H3K4me3) histone methylation marks^{17,18} (Fig. 3g, Extended Data Fig. 4a). Although active genes were associated with housekeeping functions (Extended Data Fig. 4b), *EZH1*-bound bivalent and

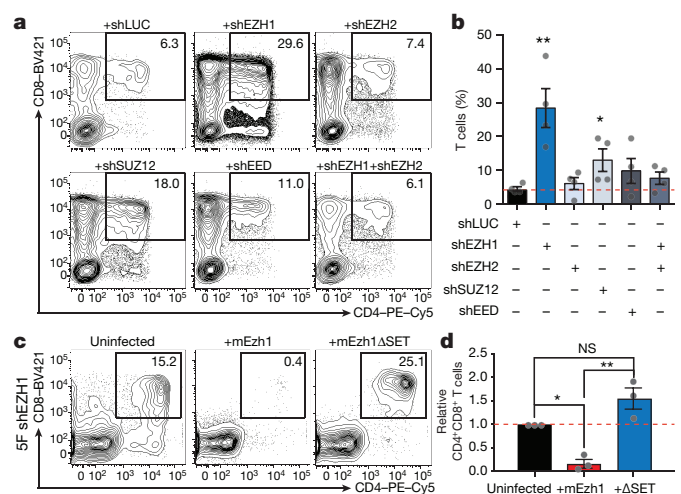


Figure 2 | Repression of canonical PRC2 subunits does not activate lymphoid potential. **a**, Representative flow plots of T cell potential of 5F cells with shRNAs targeting individual components of PRC2. **b**, Quantification of T cell potential of 5F cells plus shRNA targeting the indicated subunit in **a**, shown as using two hairpins across two independent experiments ($n = 4$). * $P = 0.0457$, ** $P = 0.0061$ by unpaired two-tailed t -test. **c**, Representative flow analysis of T cell potential in 5F cells plus shEZH1, with co-expression of full-length mouse *Ezh1* (mEzh1) or mutant mouse *Ezh1* lacking the SET domain (mEzh1ΔSET, +ΔSET). **d**, Quantification of flow analysis in **c** ($n = 3$ biological replicates). * $P = 0.0146$, ** $P = 0.0011$ by one-way ANOVA. All plots are gated on CD45⁺. Data are pooled across two independent experiments. Data are mean \pm s.e.m. NS, not significant.

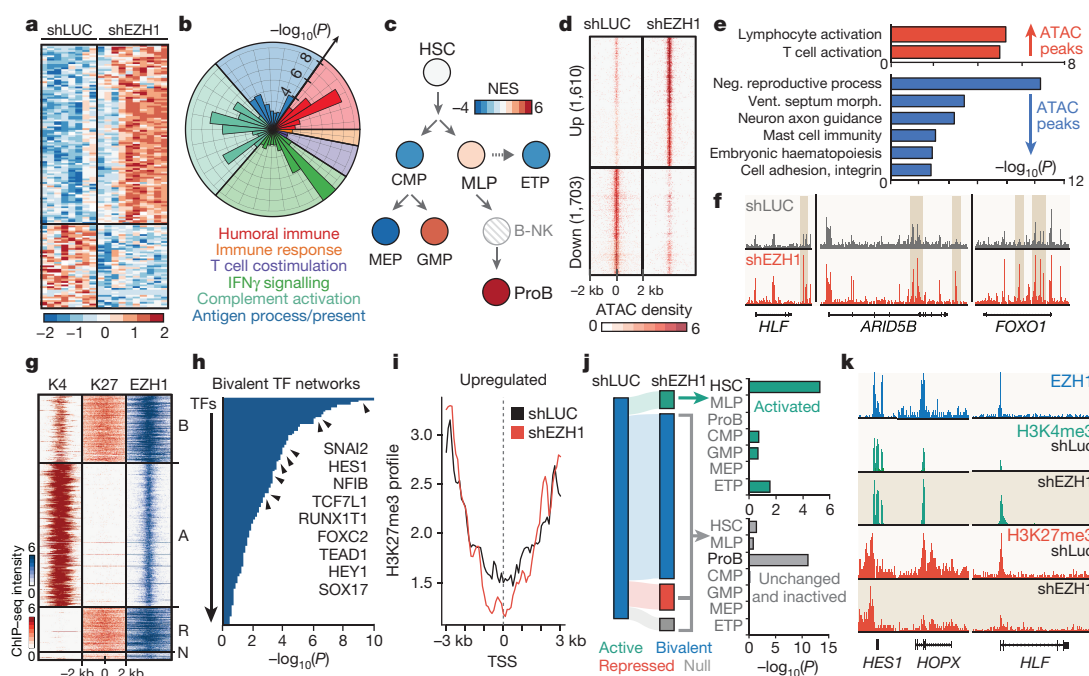


Figure 3 | EZH1 directly binds to and modulates expression and chromatin accessibility of HSC and lymphoid genes. **a**, Heat map of upregulated (104) and downregulated (49) genes (>2 -fold; Benjamini–Hochberg corrected t -test, $P < 0.1$) from RNA-seq analysis of CD34⁺CD38[−] HSPCs 5F plus shEZH1 cells ($n = 10$ biological replicates) compared to 5F plus shLUC cells ($n = 8$ biological replicates). **b**, Gene Ontology (GO) analysis of biological processes associated with significantly upregulated genes in **a**, subdivided by GO hierarchical categories with P values labelled along the radius. **c**, Enrichment of human HSC and progenitor signatures by gene set enrichment analysis (GSEA) in 5F plus shEZH1 compared with 5F plus shLUC cells, overlaid on the map of human HSPC hierarchy. CMP, common myeloid progenitor; MEP, megakaryocyte–erythroid progenitor; GMP, granulocyte–monocyte progenitor; ETP, early thymic progenitor; NES, normalized enrichment score. **d**, Density map of upregulated and downregulated ATAC peaks by MANorm²⁹ in 5F plus shEZH1 compared to 5F plus shLUC cells ($n = 2$ biological replicates). **e**, GO terms of enriched biological processes of

ATAC peaks in **d** by GREAT analysis³⁰. **f**, Tracks of representative genes that acquire a significant ATAC peak upon *EZH1* knockdown. **g**, ChIP-seq density map of EZH1 peaks within bivalent (B), repressed (R), active (A) or null (N) promoter groups ($n = 2$ biological replicates). K4, H3K4me4; K27, H3K27me3. **h**, Waterfall plot of CellNet³¹ predicted regulators of EZH1-bound bivalent gene networks. TF, transcription factor. **i**, Sitepro quantitative analysis³² of H3K27me3 levels at all upregulated genes around the transcription start site (TSS) upon *EZH1* knockdown, relative to shLUC ($n = 2$ biological replicates). **j**, Left, Sankey diagram illustrating histone methylation changes of all bivalent genes in shLUC control cells and after *EZH1* knockdown ($n = 2$ biological replicates). Right, genes that lose H3K27me3 (become activated) are specifically enriched in the HSC signature, whereas bivalent genes that are unchanged or inactivated are enriched in the ProB signature by Fisher's exact test. **k**, ChIP-seq tracks of EZH1, H3K4me3 and H3K27me3 at representative HSC promoter regions in shLUC and shEZH1 cells. Experiments replicated at least twice.

repressed genes were enriched for developmental and morphogenic processes (Extended Data Fig. 4c, d). *EZH1* knockdown increased the expression of bivalent genes, which were associated with HSC and early lymphoid lineages (Extended Data Fig. 4e, f). These genes included the targets of HSC transcription factors such as *RUNX1T1* and *SOX17*, and NOTCH factors *HES1*, *HEY1* and *FOXC2*¹⁹ (Fig. 3h). *EZH1* directly bound the promoters of HSC and ProB transcription factors including *HLF*, *PRDM16*, *LMO2*, *ETS1*, *MEIS1*, *RUNX1* and *HOX* clusters (Extended Data Fig. 4e). We also observed a global reciprocal relationship between H3K27me3 and gene transcription (Fig. 3i, Extended Data Fig. 4g–k), with poised HSC genes exhibiting loss of H3K27me3 and increased expression upon *EZH1* knockdown (Extended Data Fig. 4h, i). In total, 27 out of 29 of these activated HSC genes are direct targets of *EZH1*, including *HOPX*, *HLF*, *MEIS1* and *HES1* ($P = 7.8 \times 10^{-5}$; Fig. 3j, k).

EZH2 also bound activated HSC genes, consistent with its ability to target the same regions⁸ (Extended Data Fig. 4l); however, recent analysis of SET domain-swapping revealed context-specific sensitivity to an *EZH2*-specific inhibitor, further suggesting that although *EZH1* and *EZH2* can bind a common subset of HSC targets, these enzymes are likely to have distinct functions on chromatin²⁰. Concordant with our observation that *SUZ12* knockdown partially phenocopies *EZH1* loss (Fig. 2a, b), we observed specific enrichment of *EZH1* and *SUZ12* at activated HSC and ProB genes, consistent with non-canonical targets of the *EZH1*–*SUZ12* complex¹⁷ (Extended Data Fig. 4m–q).

Similarly, upregulated ATAC peaks in *Ezh1*-deficient AGM were also enriched for *SUZ12* binding, but not *EZH2*, indicating a conserved role for non-canonical PRC2 regulation *in vivo* (Extended Data Fig. 4r). These data suggest that in addition to the canonical function of *EZH1*–PRC2 in mediating H3K27me3 changes at poised HSC loci, *EZH1* also regulates ProB genes through a complementary non-canonical *EZH1*–*SUZ12* complex, highlighting an *EZH1*-specific function that is not phenocopied by *EZH2*.

The emergence of bona fide HSCs, defined by the capacity to repopulate irradiated adult recipients, marks the transition from embryonic to definitive haematopoiesis. We isolated AGM and yolk sac from embryonic day (E)10.5 wild-type, *Ezh1*^{+/−} and *Ezh1*^{−/−} embryos and transplanted adult non-obese diabetic (NOD)/severe combined immunodeficiency (SCID)/*Il2rg*^{−/−} (NSG) recipients (Fig. 4a). We detected peripheral blood reconstitution from wild-type AGM in 3 out of 7 mice ($11.9 \pm 7.9\%$) at 4 weeks, but chimaerism decreased by 16 weeks (2 out of 7, $12.2 \pm 8.1\%$); this corresponds to 1 repopulating unit in approximately 10.4 embryo equivalents (ee), consistent with HSCs being exceedingly rare at E10.5^{10,21}. By contrast, 5 out of 8 mice transplanted with *Ezh1*^{−/−} AGM cells were engrafted at 4 weeks ($39.2 \pm 9.4\%$) and stabilized at 16 weeks ($34.6 \pm 14.6\%$). Notably, *Ezh1*^{+/−} AGM transplant recipients had the highest initial chimaerism ($41.2 \pm 16.3\%$; 4 out of 5), which increased by 16 weeks ($68.9 \pm 17.8\%$), and was predominantly multilineage (3 out of 5) (Fig. 4e, Extended Data Fig. 5a, c). This corresponds to 1 repopulating unit in 3.6 *Ezh1*^{−/−} and

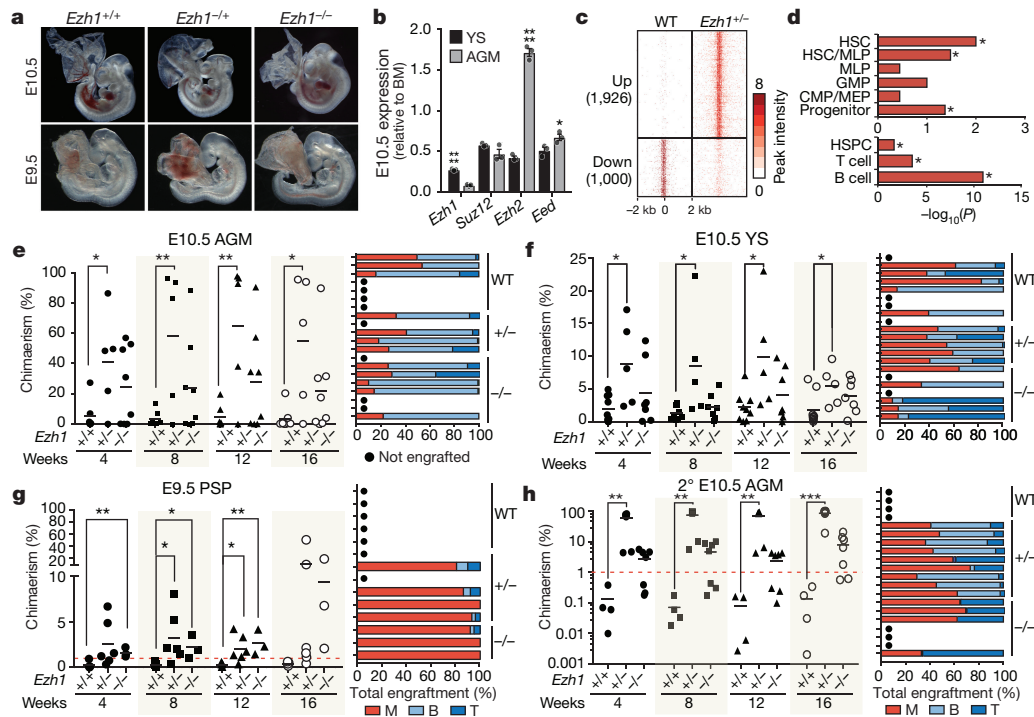


Figure 4 | *Ezh1* deficiency increases lymphoid potential and engraftment of embryonic HSPCs. **a**, Representative images of E9.5 and E10.5 embryos ($n > 50$ embryos). **b**, Quantitative PCR (qPCR) of each PRC2 subunit in E10.5 wild-type yolk sac (YS) and AGM ($n = 3$ biological replicates). $*P = 0.0439$, $****P < 0.0001$ by unpaired two-tailed t -test. Data are mean \pm s.e.m. BM, bone marrow. **c**, ATAC density map of c-Kit⁺VE-cadherin⁺CD45⁺ HSPCs sorted from 30 pooled embryos of E10.5 wild-type (WT) and *Ezh1*^{+/-} AGM. **d**, Significantly upregulated ATAC peaks were compared to HSPC, T and B cell networks and signatures of the human HSPC hierarchy⁴⁵. $*P < 0.05$ by Fisher's exact test. **e**, Left, engraftment of E10.5 AGM (3.5 ee) in sublethally irradiated adult NSG females. Donor chimaerism marked by CD45.2⁺ was measured in peripheral blood every 4 weeks up to 16 weeks post-transplantation.

2.2 *Ezh1*^{+/-} ee, or an approximately fivefold increase in HSC frequency compared to wild type.

At E10.5, the yolk sac is thought to contain few, if any, HSCs²¹. We detected low-level engraftment of wild-type yolk sac cells in 5 out of 9 recipients at 4 weeks ($3.4 \pm 0.7\%$), and in 3 out of 9 mice at 16 weeks ($4.3 \pm 1.6\%$). Most *Ezh1*^{-/-} ($4.5 \pm 0.9\%$, 6 out of 7 engrafted) and all of the *Ezh1*^{+/-} yolk-sac-transplanted mice ($5.4 \pm 1.4\%$, 5 out of 5 engrafted) showed stable long-term engraftment at 16 weeks. The number of repopulating units calculated was similar to that of the AGM (about 1 in 12.3 ee wild-type mice; 1 in 2.6 ee *Ezh1*^{-/-} mice, 1 in < 2 ee *Ezh1*^{+/-} mice). All engrafted mice were multilineage (Fig. 4f, Extended Data Fig. 5a, c). Importantly, up to 75% of peritoneal B cells in *Ezh1*^{+/-} AGM-engrafted mice were of the adult-like B-2 phenotype, as opposed to the embryonic B-1 cells (Extended Data Fig. 6a). Moreover, up to 95% of donor-derived CD45.2⁺CD3⁺ T cells expressed adult-type TCR β , as opposed to embryonic TCR $\gamma\delta$, in *Ezh1*^{-/-} and *Ezh1*^{+/-} AGM- and yolk-sac-engrafted mice (Extended Data Fig. 6b). These data provide compelling evidence that *Ezh1* deficiency, and in particular haploinsufficiency, stimulates generation of definitive HSCs and adult-like lymphopoiesis.

The para-aortic splanchnopleura (PSP) at E9.5 lacks HSCs as determined by transplantation studies³. Transplantation of E9.5 wild-type PSP cells (Fig. 4a) failed to engraft adult recipients (0 out of 5)^{21,22}; by contrast, we detected chimaerism in recipients of *Ezh1*^{-/-} (3 out of 3, $1.6 \pm 0.3\%$) and *Ezh1*^{+/-} (4 out of 6 mice, $3.6 \pm 1.3\%$) PSP at 4 weeks post-transplantation (Fig. 4g, Extended Data Fig. 5b). By 16 weeks, chimaerism increased in *Ezh1*^{-/-} (3 out of 3, $9.4 \pm 5.1\%$) and *Ezh1*^{+/-}

Each dot represents a single transplant recipient; lines denote mean values. Right, line graph distribution of engrafted mice showing T cell (T), B cell (B), and myeloid (M) contribution. **f**, Left, engraftment of E10.5 yolk sac (5 ee). Right, line graph distribution of engrafted mice. **g**, Left, engraftment of E9.5 PSP (10 ee). Right, line graph distribution of engrafted mice. **h**, Left, serial transplantation of whole bone marrow from primary recipients of E10.5 AGM cells in **e**. Secondary transplant (2°) was carried out after 24 weeks of primary transplant. Right, line graph distribution of engrafted mice ($n \geq 3$ mice per group). $*P < 0.05$, $**P < 0.01$, $***P < 0.0001$ by unpaired two-tailed t -test. See Supplementary Information for exact P values per time point. Data are pooled across four (**e–g**) or three (**h**) independent experiments; experiment in **c** was performed once.

(5 out of 6, $13.1 \pm 9.5\%$) recipients, and grafts were fully multilineage (Extended Data Fig. 5c). Thus, *Ezh1* deficiency stimulates precocious generation of bona fide HSCs during embryogenesis.

To assess the self-renewal capacity of *Ezh1*-deficient HSCs, we performed secondary transplantation. No mice showed engraftment with E10.5 wild-type AGM (0 out of 4) or yolk sac (0 out of 7). By contrast, 4 out of 7 *Ezh1*^{-/-} ($4.4 \pm 0.5\%$) and 9 out of 9 *Ezh1*^{+/-} ($57.8 \pm 10.2\%$) AGM-derived secondary recipients were engrafted (Fig. 4h, Extended Data Fig. 5d). Of note, although no *Ezh1*^{-/-} yolk sac recipients (0 out of 10) were engrafted, we observed secondary chimaerism from *Ezh1*^{+/-} yolk sac cells (5 out of 7, $1.5 \pm 0.3\%$), which increased by 16 weeks (6 out of 7, $5.1 \pm 1.9\%$) (Extended Data Fig. 5d, e). All engrafted secondary recipients were multilineage, with no evidence of leukaemic transformation (Fig. 4h, Extended Data Fig. 5c, e). Taken together, these data indicate that genetic *Ezh1* deficiency elicits precocious emergence of bona fide HSCs *in vivo*.

It has long been a curiosity that haematopoietic ontogeny progresses in reverse order, with haematopoietic progenitors appearing first in embryonic development independently of HSCs^{2,3}. We propose that *EZH1* represses definitive loci in primitive blood progenitors differentiated from human pluripotent stem cells and in mouse embryos, which precludes precocious HSC emergence during gestation. *EZH1* deficiency promotes multipotency in lineage-restricted blood progenitors and enables precocious emergence of HSCs. Although PRC2 is a well-characterized HSC regulator, our data contribute compelling evidence for the distinct molecular functions of *EZH1* and *EZH2*, and suggest a putative role for non-canonical PRC2, involving *EZH1* and

SUZ12. Homozygous loss of *Suz12* in mice impairs HSC function and lymphopoiesis, but heterozygosity for *Suz12* or *Eed* enhances HSC self-renewal^{23,24}. Consistent with this, our data reinforce the concept that HSCs are exquisitely sensitive to PRC2 dosage, with partial reduction or increase affecting function^{23–26}. Interestingly, *Runx1* haploinsufficiency also promotes premature HSC generation²⁷. Our data unify these observations; EZH1 marks many transcription factor-binding sites, whereas *Ezh1* deficiency enhances accessibility to targets of key HSC transcription factors, including *Runx1*, to promote HSC emergence (Extended Data Fig. 3k, l). We identify *Ezh1* as a molecular regulator of lineage-restricted potential of the first blood progenitors in the mammalian embryo, which accounts in part for why early embryonic progenitors lack multipotency. Beyond developmental implications, our findings suggest that resolution of EZH1-marked domains may be essential for physiological specification of HSCs from pluripotent stem cells, as a complementary approach to the synthetic reactivation of stem-cell programs by HSC transcription factors^{7,28}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 October 2016; accepted 5 December 2017.

Published online 17 January 2018.

- Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
- Medvinsky, A., Rybtsov, S. & Taoudi, S. Embryonic origin of the adult hematopoietic system: advances and questions. *Development* **138**, 1017–1031 (2011).
- Dzierzak, E. & Speck, N. A. Of lineage and legacy: the development of mammalian hematopoietic stem cells. *Nat. Immunol.* **9**, 129–136 (2008).
- McGrath, K. E. *et al.* Distinct sources of hematopoietic progenitors emerge before HSCs and provide functional blood cells in the mammalian embryo. *Cell Rep.* **11**, 1892–1904 (2015).
- Böiers, C. *et al.* Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells. *Cell Stem Cell* **13**, 535–548 (2013).
- Cedar, H. & Bergman, Y. Epigenetics of haematopoietic cell development. *Nat. Rev. Immunol.* **11**, 478–488 (2011).
- Doulatov, S. *et al.* Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. *Cell Stem Cell* **13**, 459–470 (2013).
- Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* **32**, 491–502 (2008).
- Laurenti, E. *et al.* The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat. Immunol.* **14**, 756–763 (2013).
- North, T. E. *et al.* Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity* **16**, 661–672 (2002).
- Rybtsov, S. *et al.* Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. *J. Exp. Med.* **208**, 1305–1315 (2011).
- Boisset, J. C. *et al.* *In vivo* imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature* **464**, 116–120 (2010).
- Riddell, J. *et al.* Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. *Cell* **157**, 549–564 (2014).
- Papaemmanuil, E. *et al.* Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat. Genet.* **41**, 1006–1010 (2009).
- Tothova, Z. *et al.* FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell* **128**, 325–339 (2007).
- Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
- Xu, J. *et al.* Developmental control of Polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. *Mol. Cell* **57**, 304–316 (2015).
- Margueron, R. *et al.* Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* **32**, 503–518 (2008).
- Jang, I. H. *et al.* Notch1 acts via Foxc2 to promote definitive hematopoiesis via effects on hemogenic endothelium. *Blood* **125**, 1418–1426 (2015).
- Hohmann, A. F. *et al.* Sensitivity and engineered resistance of myeloid leukemia cells to BRD9 inhibition. *Nat. Chem. Biol.* **12**, 672–679 (2016).
- Müller, A. M., Medvinsky, A., Strouboulis, J., Grosveld, F. & Dzierzak, E. Development of hematopoietic stem cell activity in the mouse embryo. *Immunity* **1**, 291–301 (1994).
- Yoder, M. C. *et al.* Characterization of definitive lymphohematopoietic stem cells in the day 9 murine yolk sac. *Immunity* **7**, 335–344 (1997).
- Lee, S. C. *et al.* Polycomb repressive complex 2 component Suz12 is required for hematopoietic stem cell function and lymphopoiesis. *Blood* **126**, 167–175 (2015).
- Xie, H. *et al.* Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. *Cell Stem Cell* **14**, 68–80 (2014).
- Majewski, I. J. *et al.* Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. *PLoS Biol.* **6**, e93 (2008).
- Kamminga, L. M. *et al.* The Polycomb group gene *Ezh2* prevents hematopoietic stem cell exhaustion. *Blood* **107**, 2170–2179 (2006).
- Cai, Z. *et al.* Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. *Immunity* **13**, 423–431 (2000).
- Sugimura, R. *et al.* Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature* **545**, 432–438 (2017).
- Shao, Z., Zhang, Y., Yuan, G. C., Orkin, S. H. & Waxman, D. J. MAnorm: a robust model for quantitative comparison of ChIP-seq data sets. *Genome Biol.* **13**, R16 (2012).
- McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
- Cahan, P. *et al.* CellNet: network biology applied to stem cell engineering. *Cell* **158**, 903–915 (2014).
- Shin, H., Liu, T., Manrai, A. K. & Liu, X. S. CEAS: cis-regulatory element annotation system. *Bioinformatics* **25**, 2605–2606 (2009).
- Doulatov, S. *et al.* Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat. Immunol.* **11**, 585–593 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Jenuwein for sharing the *Ezh1* mutant mice, which were generated at the Research Institute of Molecular Pathology (IMP, Vienna) in 2000 by D. O'Carroll (laboratory of T. Jenuwein) with the help of M. Sibilia (laboratory of E. Wagner). We also thank T. Schlaeger and the hESC Core Facility at Boston Children's Hospital for providing pluripotent stem-cell lines, R. Mathieu from BCH Flow Cytometry Core, and M. J. Chen for technical advice. This work was supported by grants from the NIH NIDDK (R24-DK092760, R24-DK49216) and NHLBI Progenitor Cell Biology Consortium (U01-HL100001); NHLBI R01HL04880 and NIH R24OD017870-01. L.T.V. is supported by the NSF Graduate Research Fellowship. M.A.K. is supported by T32 NIH Training Grant from BWH Hematology. M.C. is supported by a fellowship from the Leukemia and Lymphoma Society. S.D. is supported by K99 NIH NHLBI award (1K99HL123484). S.H.O. is an Investigator of the Howard Hughes Medical Institute. J.X. is supported by NIH grants (K01DK093543 and R01DK111430) and a Cancer Prevention and Research Institute of Texas (CPRIT) New Investigator award (RR140025). G.Q.D. was supported by the Howard Hughes Medical Institute, and is an associate member of the Broad Institute and an investigator of the Manton Center for Orphan Disease Research.

Author Contributions L.T.V., S.D. and G.Q.D. conceived the project. L.T.V. designed all experiments, performed all pluripotent stem-cell and mouse transplantation studies and interpreted data. M.A.K. analysed RNA-seq, ChIP-seq and ATAC-seq data, performed all network analyses and interpreted data. X.L. performed ChIP-seq and ATAC-seq experiments. Y.Z. and Z.S. analysed ChIP-seq and ATAC-seq data. J.B. performed and analysed qPCR and western blot validations, assisted with tissue culture, animal dissections and mouse transplantation studies. P.M.S. assisted with timed matings, animal dissections and mouse transplantation studies. D.K.J. performed western blot validations, cloned the *Ezh2*-mCherry overexpression construct, assisted with ChIP-seq optimization and interpreted data. M.C. assisted with ChIP-seq optimization. A.H. assisted with RNA-seq analysis. T.E.N., S.H.O., S.D., J.X. and G.Q.D. supervised research, interpreted data and participated in project planning. L.T.V., T.E.N., S.D. and G.Q.D. wrote the manuscript with input from all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to G.Q.D. (George.Daley@childrens.harvard.edu).

Reviewer Information Nature thanks B. Gottgens, H. Mikkola and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

A step-by-step protocol can be found at the Protocol Exchange³³.

Human iPS cell culture. All experiments were performed using MSC-iPS³⁴, CD34-iPS and CD45-iPS cells, obtained from the Boston Children's Hospital Human Embryonic Stem Cell Core (hESC) and verified by immunohistochemistry for pluripotency markers, teratoma formation and karyotyping. All cells were routinely tested for mycoplasma contamination. Human iPS cells were maintained on mouse embryonic fibroblast (GlobalStem) feeders in DMEM/F12 plus 20% KnockOut-Serum Replacement (Invitrogen), 1 mM L-glutamine, 1 mM non-essential amino acids (NEAA), 0.1 mM β -mercaptoethanol and 10 ng ml⁻¹ bFGF. Medium was changed daily, and cells were passaged 1:4 onto fresh feeders every 7 days using standard clump passing with collagenase IV.

Embryoid body differentiation. Differentiation of embryoid bodies was performed as previously described³⁵. In brief, human pluripotent stem cell colonies were scraped into non-adherent rotating 10 cm plates at the ratio of 2:1. Embryoid body medium was KO-DMEM plus 20% FBS (Stem Cell Technologies), 1 mM L-glutamine, 1 mM NEAA, 1% penicillin-streptomycin, 0.1 mM β -mercaptoethanol, 200 μ g ml⁻¹ human transferrin and 50 μ g ml⁻¹ ascorbic acid. After 24 h, medium was changed by allowing embryoid bodies to settle by gravity, and replaced with embryoid body medium supplemented with growth factors: 50 ng ml⁻¹ BMP4 (R&D Systems), 200 ng ml⁻¹ SCF, 200 ng ml⁻¹ FLT3, 50 ng ml⁻¹ G-CSF, 20 ng ml⁻¹ IL-6 and 10 ng ml⁻¹ IL-3 (all Peprotech). Medium was changed on days 5 and 10. Embryoid bodies were dissociated on day 14 by digesting with collagenase B (Roche) for 2 h, followed by treatment with enzyme-free dissociation buffer (Gibco), and filtered through an 80- μ m filter. Dissociated embryoid bodies were frozen in 10% DMSO, 40% FBS freezing solution.

Progenitor sorting. Dissociated embryoid body cells were thawed following the Lonza Poietics protocol and resuspended at 1×10^6 per 100 μ l staining buffer (PBS plus 2% FBS). CD34⁺ cells were sorted from bulk embryoid body culture using human CD34 microbeads (Miltenyi Biotec) and run through a magnetic column separator (MACS) as per the manufacturer's instructions.

Lentiviral and shRNA library plasmids. The 5F lentiviral plasmids HOXA9, ERG, RORA, SOX4 and MYB were cloned into pInducer-21 doxycycline-inducible lentiviral vector. The shRNA library targeting 20 epigenetic modifiers³⁶ was obtained from the Broad Institute RNAi Consortium in pLKO.1 or pLKO.5 lentiviral vectors. Lentiviral particles were produced by transfecting 293T-17 cells (ATCC) with the lentiviral plasmids and third-generation packaging plasmids. Viruses were harvested 24 h after transfection and concentrated by ultracentrifugation at 64,965g for 3 h using the Beckman Coulter SW 32 Ti rotor. All viruses were titred by serial dilution on 293T cells.

5F gene transfer and 5F culture. MACS-separated CD34⁺ embryoid body progenitors were seeded on retronectin-coated (10 μ g cm⁻²) 96-well plates at a density of 2×10^4 – 5×10^4 cells per well. The infection medium was SFEM (StemCell Technologies) with 50 ng ml⁻¹ SCF, 50 ng ml⁻¹ FLT3, 50 ng ml⁻¹ TPO (all R&D Systems), 50 ng ml⁻¹ IL-6 and 10 ng ml⁻¹ IL-3 (both from Peprotech). Lentiviral infections were carried out in a total volume of 150 μ l. The multiplicity of infection (MOI) for each factor was as follows: ERG MOI = 5, HOXA9 MOI = 5, RORA MOI = 3, SOX4 MOI = 3, MYB MOI = 3, and MOI = 2 for shRNA. Virus was concentrated onto cells by centrifuging the plate at 924g for 30 min at room temperature. Infections were carried out for 24 h. After gene transfer, 5F cells were cultured in SFEM with 50 ng ml⁻¹ SCF, 50 ng ml⁻¹ FLT3, 50 ng ml⁻¹ TPO, 50 ng ml⁻¹ (all R&D Systems) IL-6, and 10 ng ml⁻¹ IL-3 (Peprotech). Doxycycline (Dox) was added at 2 μ g ml⁻¹ (Sigma). Puromycin was added at 0.3 μ g ml⁻¹ (ThermoFisher Scientific). Cultures were maintained at a density of $<1 \times 10^6$ cells ml⁻¹, and the medium was changed every 3–4 days.

T cell differentiation. After 14 days of respecification, 1×10^5 5F cells were plated in OP9-DL1 stromal co-culture³⁷. Cells were cultured in α -MEM (Gibco), 1% penicillin-streptomycin, 20% FBS (Gemini), and 1 mM L-glutamine with 30 ng ml⁻¹ SCF, 5 ng ml⁻¹ FLT3, 5 ng ml⁻¹ IL-7 (all R&D Systems) for 20 days with 2 μ g ml⁻¹ Dox followed by Dox removal. Cells were collected by mechanical dissociation and filtered through a 40- μ m filter and passaged onto fresh stroma every 5–7 days. T cell development was assessed after 35 days using CD45, CD7, CD3, CD4 and CD8.

B cell differentiation. After 14 days of respecification, 5×10^4 5F cells were plated into a single well of MS-5 stroma in a 6-well NUNC plate. Cells were cultured in Myelocult H5100 (Stem Cell Technologies) supplemented with 50 ng ml⁻¹ SCF, 10 ng ml⁻¹ FLT3, 25 ng ml⁻¹ IL7, 25 ng ml⁻¹ TPO (all R&D Systems) and 1% penicillin-streptomycin for 10 days with 2 μ g ml⁻¹ Dox followed by Dox removal.

Colony assays. After 14 days of respecification, 5×10^4 cells were plated into 3 ml of complete methylcellulose H3434 (StemCell Technologies) supplemented with 10 ng ml⁻¹ IL-6 (Peprotech), 10 ng ml⁻¹ FLT3 (R&D) and 50 ng ml⁻¹ TPO (R&D)

without 2 μ g ml⁻¹ Dox. The mixture was distributed into two 60-mm dishes and maintained in a humidified chamber for 14 days.

Mouse transplantation. NOD/SCID/IL2rg^{-/-} (NSG) (Jackson Laboratory) mice were bred and housed at the Boston Children's Hospital animal care facility. Animal experiments were performed in accordance with institutional guidelines approved by Boston Children's Hospital Animal Care Committee. At least three animals were used per cohort, based on previous transplantation studies. Mice were assigned randomly to groups and blinding was not used. In brief, 8–12-week-old mice were irradiated (2.75 Gy) 24 h before transplant. To ensure consistency between experiments, only female mice were used. Sublethally irradiated adult NSG females were transplanted intravenously with 3.5 ee of whole E10.5 AGM, 5 ee of whole E10.5 yolk sac or 10 ee of whole E9.5 PSP. Mice were bled retroorbitally every 4 weeks to monitor donor chimaerism up to 16 weeks post-transplantation. Twenty-four weeks after primary transplantation, primary recipients from each group were euthanized and 4×10^6 whole bone marrow cells were transplanted into 1–3 secondary recipients. Cells were transplanted in a 200 μ l volume using a 28.5-gauge insulin needle. Sulfatrim was administered in drinking water to prevent infections after irradiation. Data points were combined from all independent experiments and outliers were not excluded.

Flow cytometry. The following antibodies were used for human cells: CD45 allophycocyanin (APC)-conjugated Cy7 (557833, BD Biosciences), CD4 phycoerythrin (PE)-conjugated Cy5 (IM2636U, Beckman Coulter Immunotech), CD8-BV421 (RPA-T8, BD Horizon), CD5-BV510 (UCHT2, BD Biosciences), TCR $\gamma\delta$ -APC (555718, BD Biosciences), TCR $\alpha\beta$ -BV510 (T10B9.1A-31, BD Biosciences), CD3-PE-Cy7 (UCHT1, BD Pharmingen), CD7-PE (555361, BD Pharmingen), CD1a-APC (559775, BD Pharmingen) for T cell staining. For B cell staining: CD45-PE-Cy5 (IM2652U, Beckman Coulter Immunotech), CD19-PE (4G7, BD Biosciences), CD56-V450 (B159, BD Biosciences), CD11b-APC-Cy7 (557754, BD Biosciences), For HSC/progenitor sorting: CD34-PE-Cy7 (8G12, BD Biosciences), CD45-APC-Cy7 (557833, BD Biosciences), CD38-PE-Cy5 (IM2651U, Beckman Coulter) and DAPI. For myeloid and erythroid staining: CD11b-APC-Cy7 (557754, BD Biosciences), GLYA-PE-Cy7 (A71564, Beckman Coulter), CD71-PE (555537, BD Biosciences), CD45-PE-Cy5 (IM2652U, Beckman Coulter Immunotech). All staining was performed with $<1 \times 10^6$ cells per 100 μ l staining buffer (PBS plus 2% FBS), with a 1:100 dilution of each antibody, for 30 min at room temperature in the dark. Compensation was performed by automated compensation with anti-mouse IgG and negative beads (BD Biosciences). All acquisitions were performed on a BD Fortessa or BD Aria cytometer.

The following antibodies were used for mouse cells: CD45.2-PE-Cy7 (104, eBioscience), CD45.1-FITC (A20, eBioscience), B220-PB (RA3-6B2, BD Biosciences), Ter119-PE-Cy5 (Ter 119, eBioscience), GR1 (RB6-8C5, BD Bioscience), CD3-APC (145-2C11, eBioscience), CD19-APC-Cy7 (1D3, BD Bioscience), MAC1-AF700 (M1/70, BD Bioscience) for engraftment analyses. For B cell staining: CD45.2-APC-Cy7 (104, BioLegend), CD23-PE-Cy7 (B3B4, eBioscience), Ter119-PE-Cy5 (Ter 119, eBioscience), MAC1-A700 (M1/70, BD Bioscience), CD5-BV510 (53-7.3, BD Biosciences), IgM-eFluor660 (II/41, eBioscience). For T cell staining: CD45.2-PE-Cy7 (104, eBiosciences), TCR β -PE-Cy5 (H57-597, BD Biosciences), CD8-APC-EF780 (53-6.7, eBioscience), CD4-APC (GK1.5, eBioscience), CD3-AF700 (17A2, BioLegend), TCR $\gamma\delta$ -FITC (GL3, BD Biosciences). For HSPC sorting: CD16/32 (93, BioLegend), Ter119-biotin (Ter119, eBioscience), Gr-1-biotin (RB6-8C5, eBioscience), CD3-biotin (17A2, eBioscience), CD5-biotin (53-7.3, eBioscience), CD8-biotin (53-6.7, eBioscience), CD19-biotin (eBio1D3, eBioscience), streptavidin-eFluor450 (eBioscience), CD45-PerCP-Cy5.5 (30-F11, eBioscience), CD144-eFluor660 (eBioBV13, eBioscience), CD117-APC-eFluor 780 (2B8, eBioscience), CD41-PE-Cy7 (eBioMWReg30, eBioscience). All staining was performed with $<1 \times 10^6$ cells per 100 μ l staining buffer (PBS plus 2% FBS), with a 1:100 dilution of each antibody, for 30 min on ice in the dark. Compensation was performed by automated compensation with anti-rat and anti-hamster IgG and negative beads (BD Biosciences). All acquisitions were performed on a BD Fortessa or BD Aria cytometer.

RNA-seq. Human cells were stained and sorted using CD34-PE-Cy7 (8G12, BD Biosciences), CD38-PE-Cy5 (IM2651U, Beckman Coulter) and DAPI (Beckman Coulter). RNA-seq libraries were prepared using the NEB Ultra (PolyA) kit as per the manufacturer's protocol with 50 ng input RNA. Mouse cells were stained and sorted using the 'HSPC stain' (see 'Flow cytometry'). RNA-seq libraries were prepared using the Clontech SMARTer Universal Low Input kit as per the manufacturer's protocol with 10 ng input RNA. Libraries were sequenced using the 200 cycle paired-end kit on the Illumina HiSeq2500 system. RNA-seq reads were analysed with Tuxedo Tools following a standard protocol³⁸. Reads were mapped with TopHat version 2.1.0 and Bowtie2 version 2.2.4 with default parameters against build hg19 of the human genome, and build hg19 of the RefSeq human genome

annotation. Samples were quantified with the Cufflinks package version 2.2.1. Differential expression was performed using Cuffdiff with default parameters.

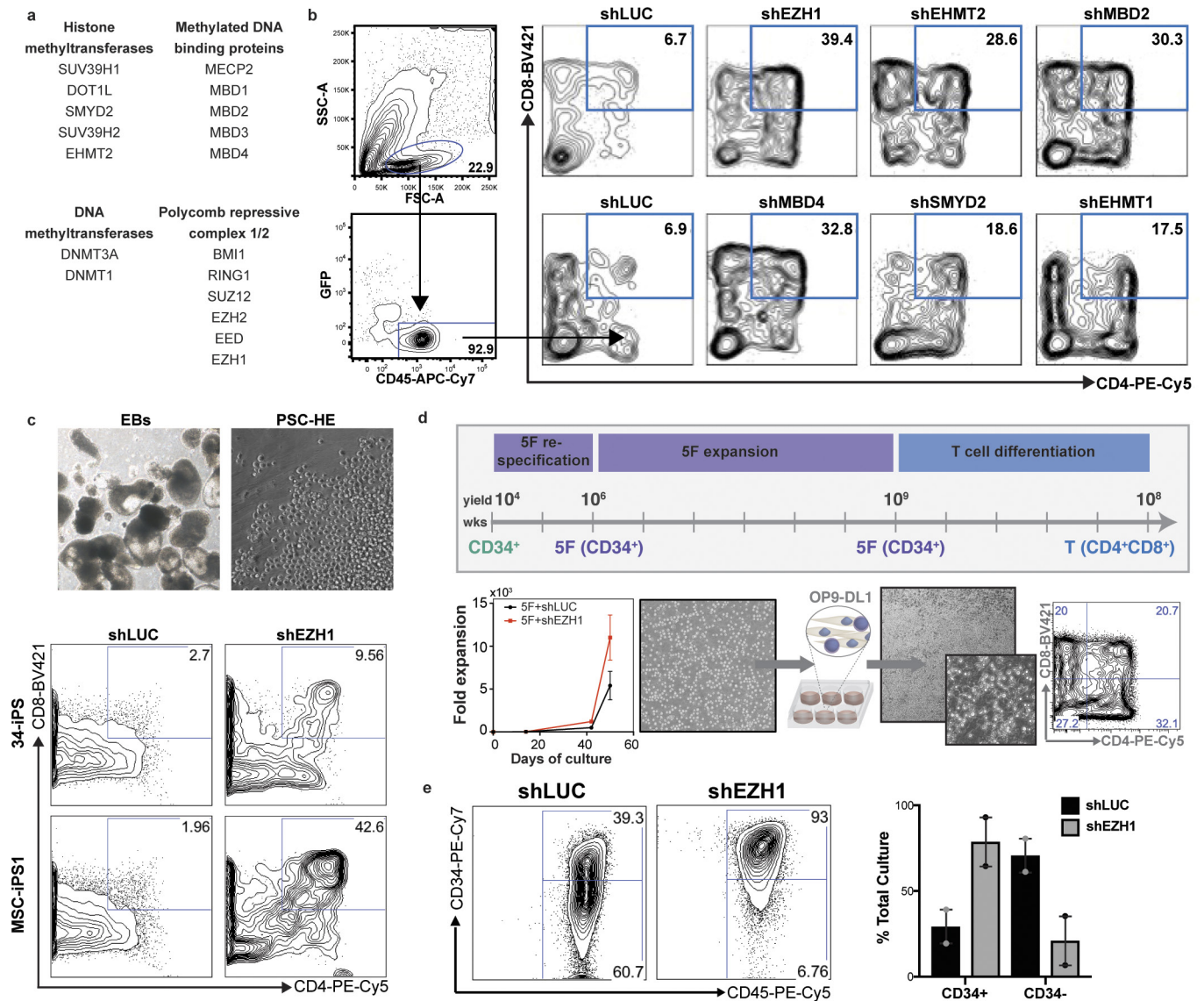
ATAC-seq. ATAC-seq was performed as previously described³⁹. 5×10^3 – 50×10^3 cells were used for each tagmentation using Tn5 transposases. The resulting DNA was isolated, quantified and sequenced on an Illumina NextSeq500 system. The raw reads were aligned to the human genome assembly hg19 using Bowtie⁴⁰ with the default parameters, and only tags that uniquely mapped to the genome were used for further analysis. ATAC peaks were identified using MACS⁴¹.

ChIP-seq. ChIP experiments were performed as previously described⁴² using the antibodies for H3K4me3 (04-745, Millipore) and H3K27me3 (07-449, Millipore) in 5F cells. For bioChIP analysis of EZH1 or EZH2 occupancy, Flag-biotin-tagged EZH1 or EZH2 was stably expressed in 5F cells. The chromatin was isolated and immunoprecipitated with streptavidin Dynabeads (Life Technologies) as previously described⁴³. ChIP-seq libraries were generated using NEBNext ChIP-seq Library Prep Master Mix following the manufacturer's protocol (New England Biolabs), and sequenced on an Illumina NextSeq500 system. ChIP-seq raw reads were aligned to the human genome assembly hg19 using Bowtie⁴⁰ with the default parameters; only tags that uniquely mapped to the genome were used for further analysis. ChIP-seq peaks were identified using MACS⁴¹.

Bioinformatics and statistical analysis. All statistical calculations were performed using GraphPad Prism. Tests between two groups used a two-tailed unpaired Student's *t*-test. Data are presented as mean \pm s.e.m. Where indicated, ANOVA was used, with $P < 0.05$ considered significant. GSEA and GO were run according to default parameters in their native implementations. Statistical enrichment of gene lists was performed using Fisher's exact test. No statistical methods were used to predetermine sample size.

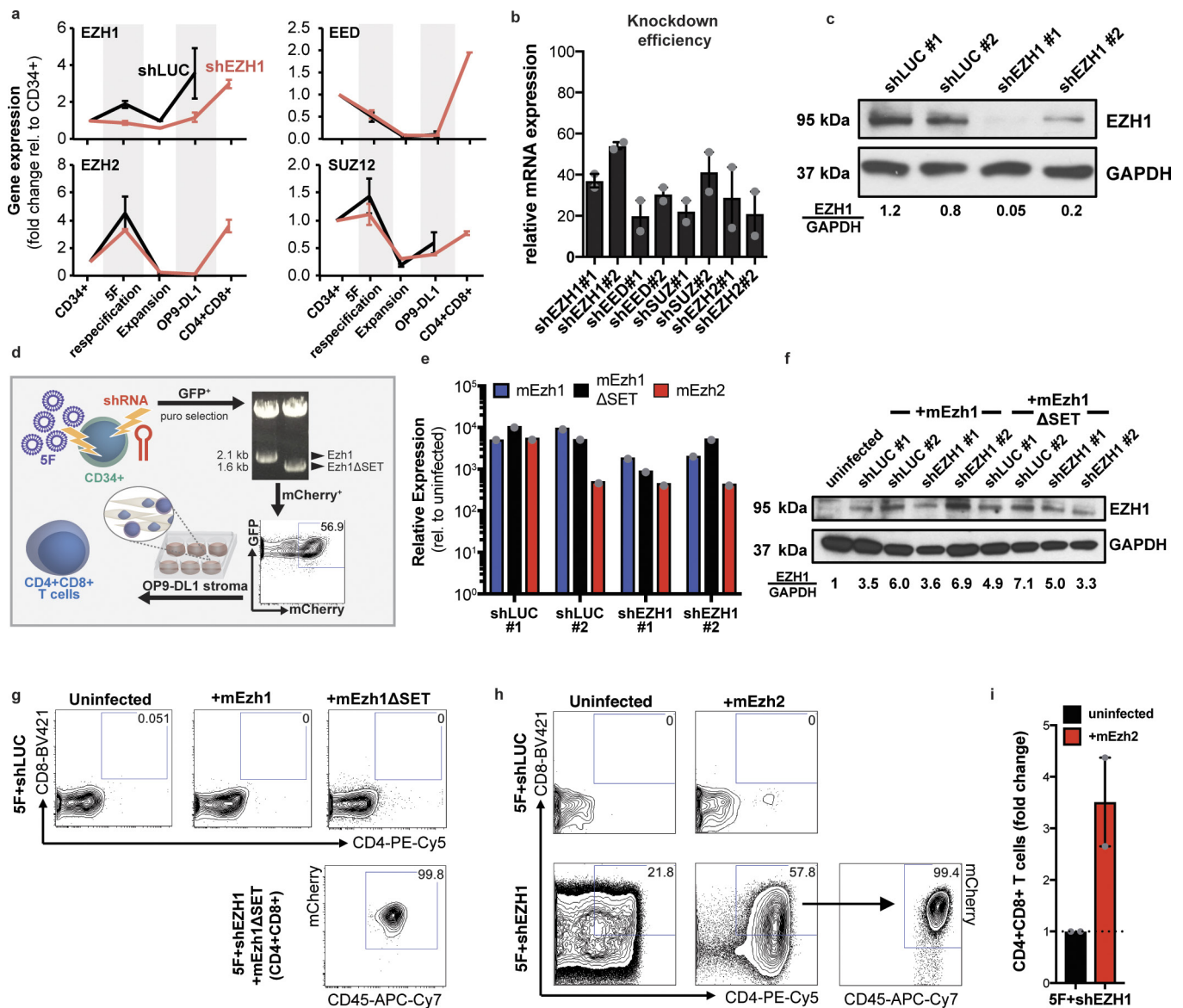
Data availability. All RNA-seq, ATAC-seq and ChIP-seq data have been deposited to the Gene Expression Omnibus (GEO) database under the accession number GSE89418.

33. Vo, L. T., Doulatov, S. & Daley, G. Q. Haematopoietic progenitor and lymphoid differentiation from human pluripotent stem cells. *Protoc. Exch.* <http://doi.org/10.1038/protex.2017.160> (2018).
34. Park, I. H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141–146 (2008).
35. Chadwick, K. *et al.* Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. *Blood* **102**, 906–915 (2003).
36. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483**, 598–602 (2012).
37. Holmes, R. & Zúñiga-Pflücker, J. C. The OP9-DL1 system: generation of T-lymphocytes from embryonic or hematopoietic stem cells *in vitro*. *Cold Spring Harb. Protoc.* **2009**, <http://doi.org/10.1101/pdb.prot5156> (2009).
38. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
39. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
40. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
41. Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
42. Huang, J. *et al.* Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. *Dev. Cell* **36**, 9–23 (2016).
43. Kim, J., Cantor, A. B., Orkin, S. H. & Wang, J. Use of *in vivo* biotinylation to study protein–protein and protein–DNA interactions in mouse embryonic stem cells. *Nat. Protoc.* **4**, 506–517 (2009).
44. Ditadi, A. *et al.* Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nat. Cell Biol.* **17**, 580–591 (2015).



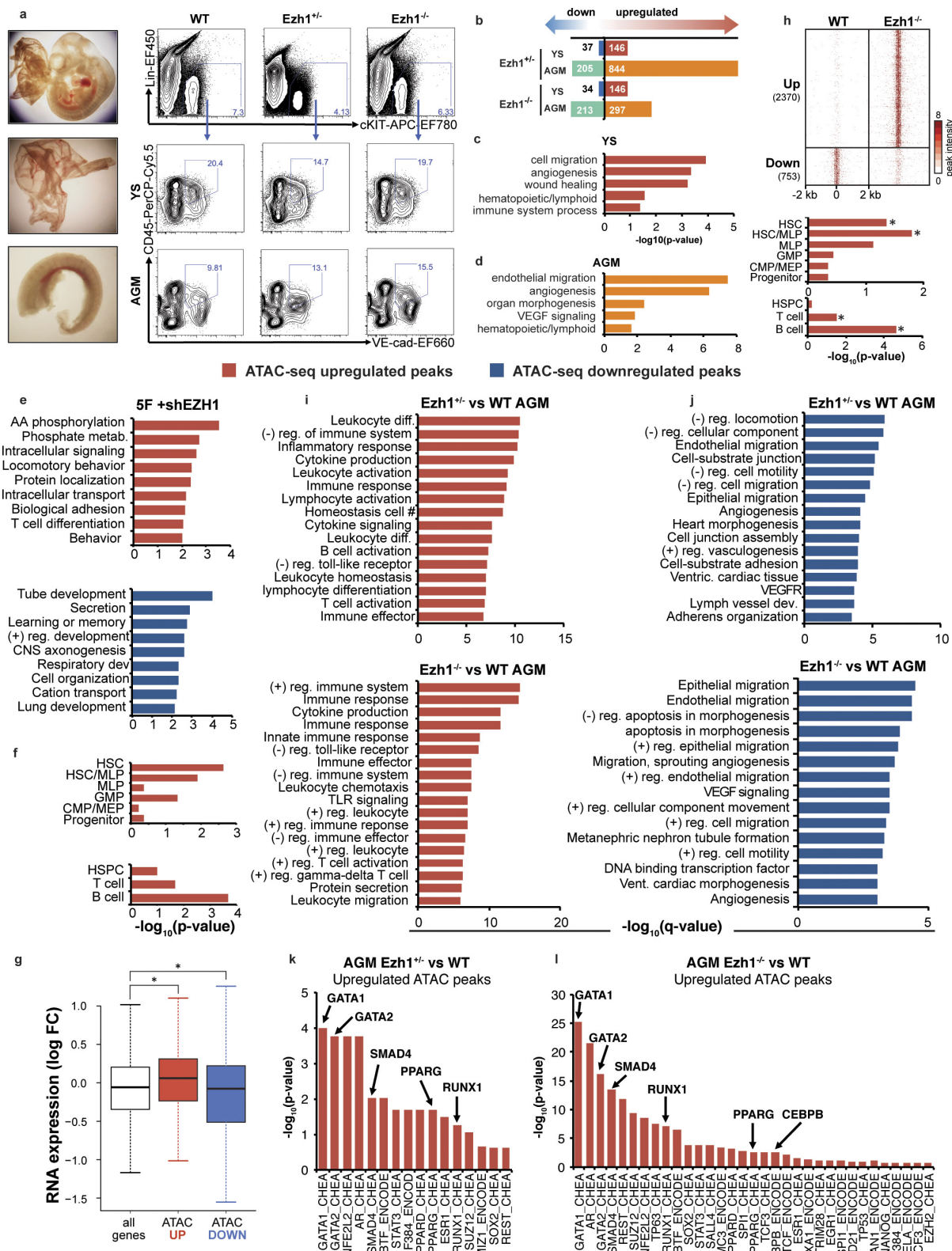
Extended Data Figure 1 | *EZH1* knockdown activates lymphoid potential from pluripotent stem cells. **a**, List of all candidate epigenetic modifiers in loss-of-function shRNA screen. **b**, Representative flow plots of $CD4^+CD8^+$ T cell potential across top six candidates from four independent hairpins in two independent experiments ($n = 8$). See Fig. 1. **c**, $CD34^+$ cells were isolated after 9 days of embryoid body (EB) differentiation (top left), transduced with shLUC or shEZH1 and cultured under conditions that promote endothelial-to-haematopoietic transition⁴⁴. After 6 days, rounded haematopoietic cells (top right) were collected and co-cultured on OP9-DL1 stroma. Bottom, flow cytometric analysis of T cell potential in shLUC and shEZH1 cells without 5F is shown for two independent iPS lines (34-iPS and MSC-iPS1) in one experiment ($n = 2$ biological replicates). PSC-HE, pluripotent stem-cell-derived haemogenic

endothelium. **d**, Expansion and differentiation potential of 5F plus shEZH1 cells after long-term *in vitro* culture. 5F plus shEZH1 cells were maintained in cultures containing doxycycline for 14 days respecification (approximately 100-fold expansion), plus an additional 6 weeks (approximately 1,000-fold expansion) and then plated into OP9-DL1 stromal cells for T cell differentiation. Representative flow cytometric analyses of T cell potential of 5F plus shLUC and 5F plus shEZH1 cells after 13 weeks of expansion and differentiation ($n = 2$ biological replicates). **e**, Flow cytometric analysis (left) and quantification (right) of the proportion of $CD34^+$ and $CD34^-$ haematopoietic progenitors in doxycycline-containing suspension culture at day 25 ($n = 2$ biological replicates).



Extended Data Figure 2 | Ezh1, but not Ezh2, suppresses T cell potential and requires its catalytic domain. **a**, qPCR of PRC2 expression (human genes *EZH1*, *EED*, *EZH2* and *SUZ12*), during the course of differentiation from human pluripotent stem cell-derived CD34⁺ cells, respecification (5F), expansion, OP9-DL1 co-culture and CD4⁺CD8⁺ T cells ($n = 2$ biological replicates in one experiment). **b**, qPCR of mRNA knockdown efficiency of individual shRNAs for PRC2 genes ($n = 2$ biological replicates). See also Fig. 2a, b, c, Western blot for EZH1 and GAPDH protein levels. **d**, Scheme for rescue experiments. GFP⁺ 5F cells were transduced with shRNAs and selected with puromycin. 5F plus shRNA cells were then transduced with full-length mouse *Ezh1* open-reading frame (*mEzh1*) or mutant mouse *Ezh1* with the catalytic SET domain deleted (*mEzh1*ΔSET), marked by mCherry fluorescence. Triple-transduced (GFP⁺, puromycin-resistant, mCherry⁺) cells were sorted and seeded onto OP9-DL1. T cells were analysed by flow cytometry

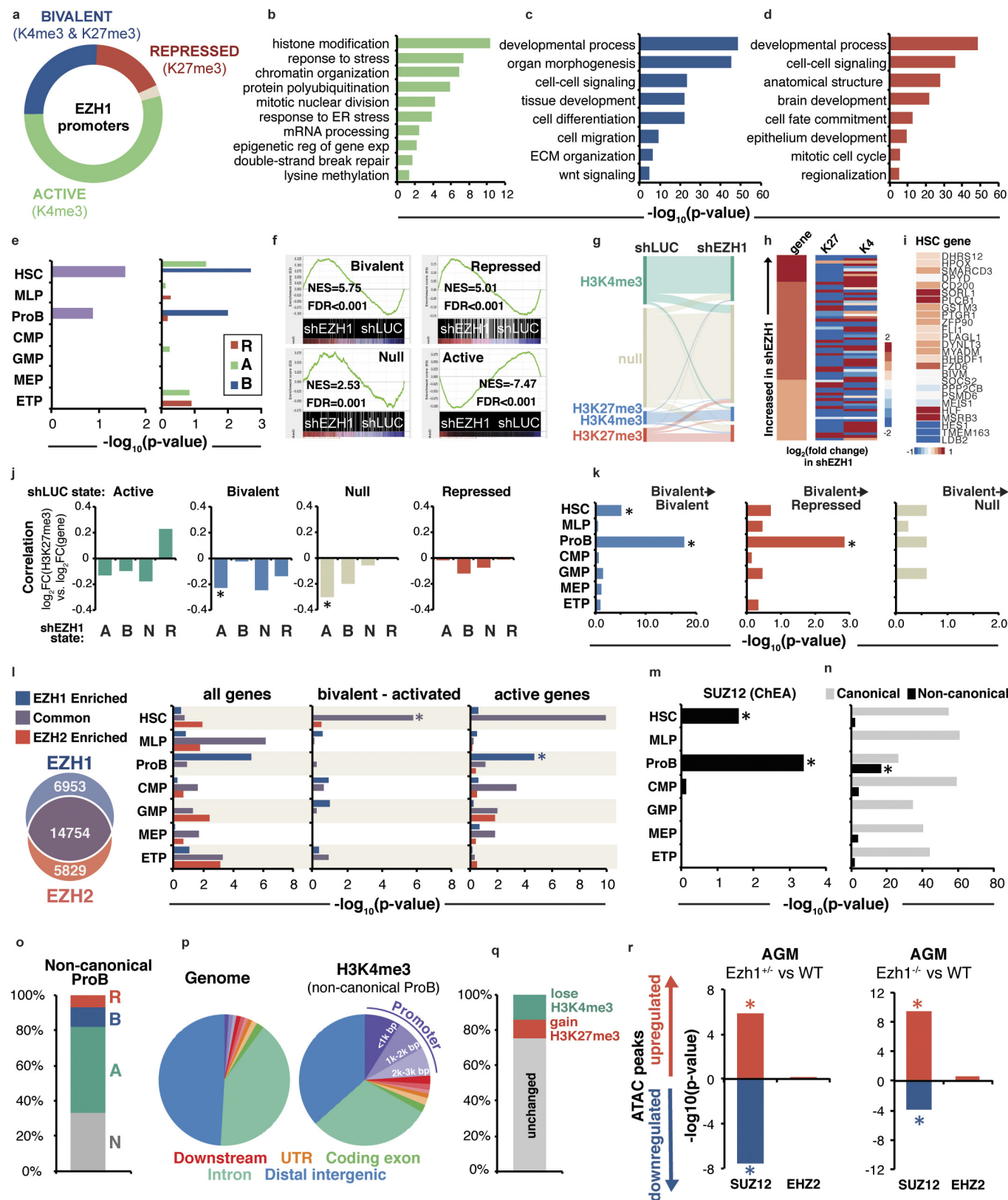
after 5 weeks of differentiation. See also Fig. 2c. **e**, Expression of full-length mouse *Ezh1*, catalytic-deleted *mEzh1*ΔSET, or full-length mouse *Ezh2* in shLUC and shEZH1 cells by qPCR. **f**, Western blot validation of expression of mouse *Ezh1* or mutant *mEzh1*ΔSET in shLUC and shEZH1 cells. **g**, Top, representative flow cytometry plots of T cell potential for 5F plus shLUC cells for rescue experiments in **d** ($n = 3$ biological replicates). Bottom, CD4⁺CD8⁺ T cells were verified for mCherry fluorescence. See also Fig. 2c. All plots are gated on CD45⁺. **h**, 5F plus shRNA cells were transduced with full-length mouse *Ezh2* open-reading frame (*mEzh2*) marked by mCherry fluorescence. Triple-transduced (GFP⁺, puromycin-resistant, mCherry⁺) cells were sorted and seeded onto OP9-DL1 stromal cells. T cells were analysed by flow cytometry after 5 weeks of differentiation. Representative flow plots for two biological replicates in one experiment. **i**, Quantification of data in **h**. Data are mean \pm s.e.m.



Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | *Ezh1* regulates haematopoietic and lymphoid programs *in vitro* and *in vivo*. **a**, Representative images of E10.5 embryo (top), yolk sac (middle) and AGM (bottom) from $n > 30$ embryos. Lin⁻c-Kit⁺VE-cadherin⁺CD45⁺CD41⁺ cells from E10.5 yolk sac and AGM were FACS-sorted followed by RNA-seq analysis. See also Fig. 4a, c, d. **b**, Genes upregulated and downregulated by more than twofold in *Ezh1*^{+/-} or *Ezh1*^{-/-} yolk sac and AGM compared to those from wild-type mice. **c**, **d**, GO term annotations of upregulated genes in *Ezh1*^{+/-} and *Ezh1*^{-/-} yolk sac and AGM compared to those from wild-type mice. **e**, GO analysis of enriched pathways of the 1,033 nearest neighbour genes associated with upregulated ATAC peaks (top) and the nearest 1,012 neighbour genes associated with downregulated ATAC peaks (bottom). See also Fig. 3d, e. **f**, Comparison of upregulated ATAC peaks in 5F plus shEZH1 cells with HSPC hierarchy signatures⁴⁵ (top) and HSPC B and T cell networks (bottom). See also Fig. 3d, e. **g**, Box plot of expression of genes associated

with upregulated and downregulated ATAC peaks. * $P < 0.05$ by one-way ANOVA. **h**, ATAC density map of c-Kit⁺VE-cadherin⁺CD45⁺ HSPCs sorted from approximately 30 embryos of E10.5 wild-type and *Ezh1*^{-/-} AGM (top) from one experiment. Significantly upregulated ATAC peaks were compared to HSPC, T, B cell networks and signatures of the human HSPC hierarchy (bottom). See also Fig. 4c, d. **i**, GO terms of enriched pathways of regions associated with significantly upregulated ATAC peaks annotated by GREAT analysis in *Ezh1*^{+/-} AGM (top) and *Ezh1*^{-/-} AGM (bottom) compared to wild type. See also Fig. 4c, d. **j**, GO terms of enriched pathways of regions associated with significantly downregulated ATAC peaks annotated by GREAT analysis³⁰ in *Ezh1*^{+/-} AGM (top) and *Ezh1*^{-/-} AGM (bottom) compared to wild type. See also Fig. 4c. **k**, Transcription factor binding to genes with upregulated ATAC peaks in *Ezh1*^{+/-} (left) and *Ezh1*^{-/-} (right) AGM from **i** compared to wild-type AGM.

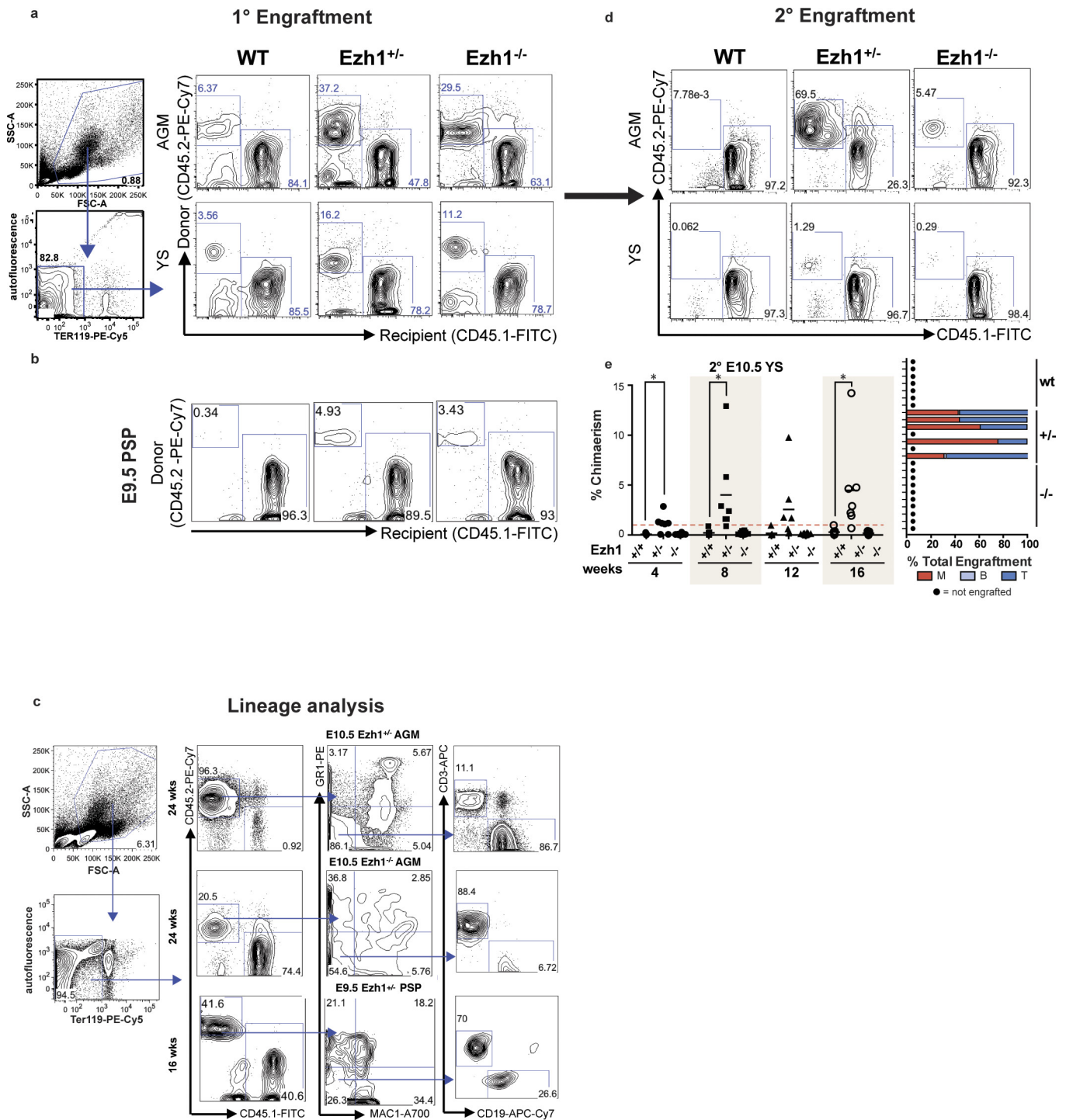


Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Genome-wide chromatin occupancy reveals EZH1 enrichment at bivalent HSC genes and non-canonical active lymphoid genes.

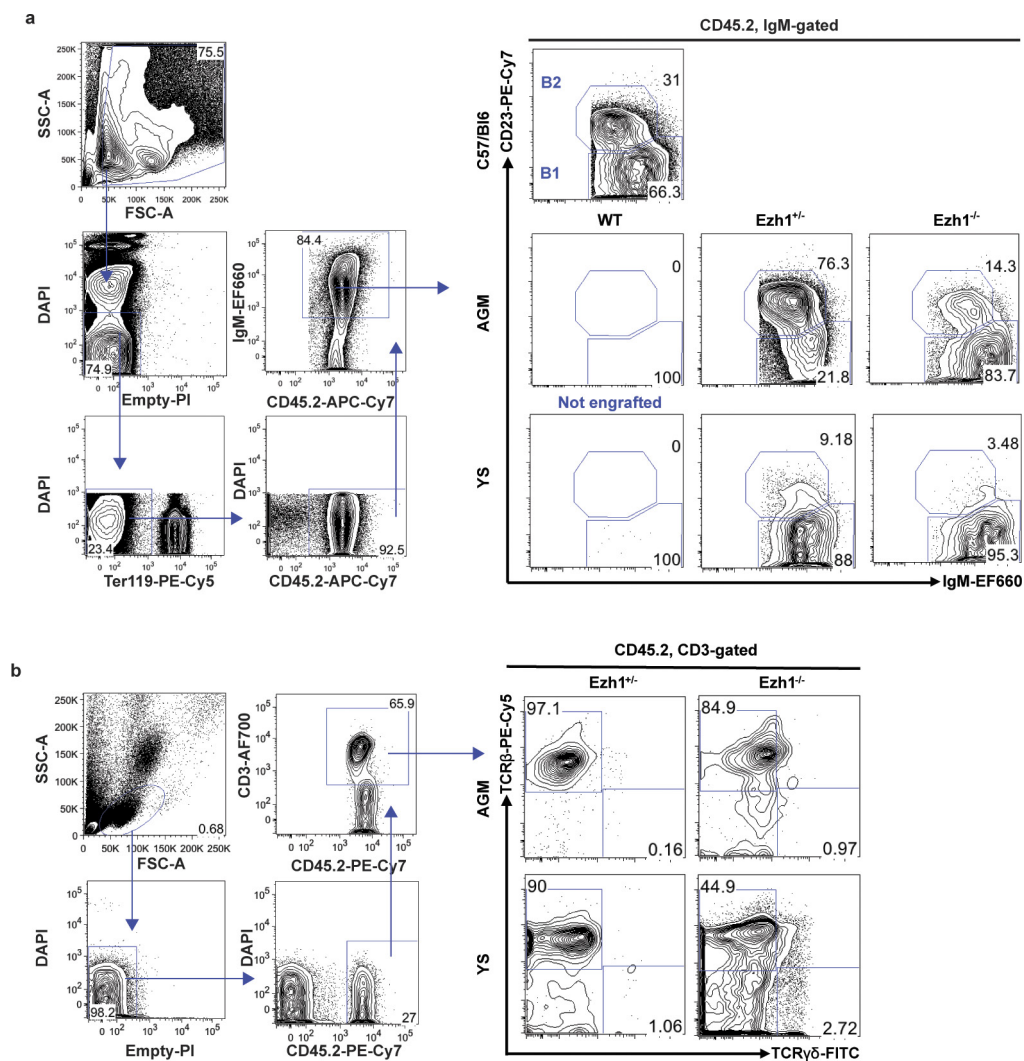
a, Breakdown of EZH1 binding at promoter regions and associated histone marks. **b–d**, GO term analysis of EZH1-bound active (**b**), bivalent (**c**) and repressed (**d**) genes. **e**, Distribution of EZH1-bound genes across the haematopoietic hierarchy (left) and their associated histone marks (right). A, active (H3K4me3-marked); B, bivalent (H3K4me3 and H3K27me3-marked); R, repressed (H3K27me3-marked). **f**, GSEA analysis of EZH1-bound genes correlated with RNA-seq data upon *EZH1* knockdown. **g**, Sankey diagram showing genome-wide changes in histone methylation status upon *EZH1* knockdown. **h**, Upregulated genes exhibit reciprocal decreases in H3K27me3 levels, as quantified by EpiChIP software. K4, H3K4me3; K27, H3K27me3. See also Fig. 3i. **i**, Activated (formerly bivalent) HSC genes exhibit increased gene expression upon *EZH1* knockdown and loss of H3K27me3. See also Fig. 3j. **j**, Correlations between changes in H3K27me3 and gene expression

levels upon *EZH1* knockdown, subdivided by subgroups corresponding to methylation changes. N, null. **k**, Breakdown of bivalent–bivalent (left), bivalent–repressed (centre) and bivalent–null (right) genes upon *EZH1* knockdown across the haematopoietic hierarchy. **l**, Overlap of EZH1- and EZH2-enriched peaks and the distribution of all EZH1-enriched, EZH2-enriched or common genes across the hierarchy (left), or specifically bivalent genes that become activated upon *EZH1* knockdown (middle) and active genes, marked by H3K4me3 in shLUC (right). **m**, *SUZ12* binding (from the ChEA database) across the haematopoietic hierarchy. **n**, Canonical and non-canonical previously identified targets¹⁷ across the haematopoietic hierarchy. **o**, **p**, Breakdown of histone marks on non-canonical ProB genes (**o**) and the genome-wide distribution from CEAS analysis³² (**p**). **q**, Changes in actively marked, non-canonical ProB genes (green bar in **o**), upon *EZH1* knockdown. **r**, *SUZ12* and *EZH2* binding (ChEA database) at ATAC peaks in *Ezh1*^{+/-} and *Ezh1*^{-/-} AGM. **P* < 0.05 by one-way ANOVA.



Extended Data Figure 5 | *Ezh1* deficiency enhances embryonic HSPC engraftment. **a**, Whole E10.5 AGM and yolk sac were transplanted intravenously into sublethally irradiated NSG adult females. Chimaerism was monitored by retroorbital bleeding every 4 weeks. Representative flow plots are shown for analysis after 4 weeks in $n \geq 3$ mice. See also Fig. 4e, f. **b**, Whole E9.5 PSP was transplanted intravenously into sublethally irradiated NSG adult females. Chimaerism was monitored via retroorbital bleeding every 4 weeks. Representative flow plots are shown for analysis after 8 weeks in $n \geq 3$ mice. See also Fig. 4g. **c**, Representative flow plots of lineage analysis in E10.5 AGM *Ezh1*^{+/-} and *Ezh1*^{-/-} primary transplant recipients after 24 weeks, and in E9.5 PSP *Ezh1*^{+/-} primary

transplant recipient after 16 weeks ($n \geq 3$ mice per group). See also Fig. 4e, g. **d**, Primary recipients in **a** were euthanized 24 weeks post-transplantation and 4×10^6 whole bone marrow was transplanted into sublethally irradiated adult NSG females intravenously. Chimaerism was monitored via retroorbital bleeding. Representative flow plots of E10.5 AGM and yolk sac secondary transplants after 4 weeks in $n \geq 3$ mice. See also Fig. 4h. **e**, Left, secondary transplantation of E10.5 yolk sac primary recipients (Fig. 4f). Right, lineage distribution of E10.5 secondary recipients. Data are pooled across three independent experiments. * $P < 0.05$, ** $P < 0.01$ by unpaired two-sided *t*-test; see Supplementary Information for exact *P* values per time point.



Extended Data Figure 6 | *Ezh1*-deficient embryonic HSPCs contribute to adult-type lymphopoiesis *in vivo*. **a**, Flow analysis of B1 and B2 progenitors in the peritoneal cavity of engrafted primary recipients ($n = 1$

mouse per group). **b**, Flow analysis of TCR β and TCR $\gamma\delta$ frequencies of donor-derived peripheral CD3⁺ T cells from engrafted primary recipients ($n = 1$ mouse per group). See also Fig. 4e, f.

Clonal evolution mechanisms in *NT5C2* mutant-relapsed acute lymphoblastic leukaemia

Gannie Tzoneva^{1†*}, Chelsea L. Dieck^{1*}, Koichi Oshima¹, Alberto Ambesi-Impiombato^{1†}, Marta Sánchez-Martín¹, Chioma J. Madubata², Hossein Khiabanian³, Jiangyan Yu^{4,5}, Esme Waanders⁴, Ilaria Iacobucci⁶, Maria Luisa Sulis⁷, Motohiro Kato⁸, Katsuyoshi Koh⁸, Maddalena Paganin⁹, Giuseppe Basso⁹, Julie M. Gastier-Foster^{10,11,12,13}, Mignon L. Loh^{14,15}, Renate Kirschner-Schwabe¹⁶, Charles G. Mullighan⁶, Raul Rabadan^{2,17} & Adolfo A. Ferrando^{1,2,7,18}

Relapsed acute lymphoblastic leukaemia (ALL) is associated with resistance to chemotherapy and poor prognosis¹. Gain-of-function mutations in the 5'-nucleotidase, cytosolic II (*NT5C2*) gene induce resistance to 6-mercaptopurine and are selectively present in relapsed ALL^{2,3}. Yet, the mechanisms involved in *NT5C2* mutation-driven clonal evolution during the initiation of leukaemia, disease progression and relapse remain unknown. Here we use a conditional-and-inducible leukaemia model to demonstrate that expression of *NT5C2*(R367Q), a highly prevalent relapsed-ALL *NT5C2* mutation, induces resistance to chemotherapy with 6-mercaptopurine at the cost of impaired leukaemia cell growth and leukaemia-initiating cell activity. The loss-of-fitness phenotype of *NT5C2*^{+/R367Q} mutant cells is associated with excess export of purines to the extracellular space and depletion of the intracellular purine-nucleotide pool. Consequently, blocking guanosine synthesis by inhibition of inosine-5'-monophosphate dehydrogenase (IMPDH) induced increased cytotoxicity against *NT5C2*-mutant leukaemia lymphoblasts. These results identify the fitness cost of *NT5C2* mutation and resistance to chemotherapy as key evolutionary drivers that shape clonal evolution in relapsed ALL and support a role for IMPDH inhibition in the treatment of ALL.

Improved support and intensified chemotherapy regimens have increased the overall survival rates of newly diagnosed paediatric ALL to over 80%¹. However, the outcomes of patients with relapsed or refractory ALL remain poor, with cure rates of only about 40%¹. Leukaemia-initiating cells capable of self-renewal^{4,5}, protective microenvironment safe-haven niches^{6,7} and clonal evolution^{8–10} with acquisition of secondary genetic alterations driving chemotherapy resistance^{2,3,9–13} have all been implicated as drivers of ALL disease progression and relapse. In this context, heterozygous activating mutations in the *NT5C2* nucleotidase gene are present in about 20% of relapsed paediatric T-cell ALL (T-ALL) cases² and 3–10% of relapsed B-precursor ALLs^{2,3}. *NT5C2* (Enzyme Commission (EC) number 3.1.3.5) is a highly conserved and ubiquitously expressed enzyme responsible for catalysing the 5'-dephosphorylation of the purine nucleotides inosine monophosphate, xanthine monophosphate and guanosine monophosphate¹⁴. This activity controls the intracellular levels of 6-hydroxypurine monophosphate nucleotides via their dephosphorylation to nucleosides, which are subsequently exported

out of the cell^{14,15}. In addition, *NT5C2* metabolizes and inactivates the active metabolites that mediate the cytotoxic activity of 6-mercaptopurine (6-MP), a purine analogue chemotherapy drug that is broadly used in the treatment of ALL¹⁶ (Extended Data Fig. 1). Expression of gain-of-function relapse-associated mutant forms of *NT5C2* can therefore induce resistance to 6-MP *in vitro*^{2,3}.

Genomic profiling of matched samples from the time of ALL diagnosis and after relapse supports the hypothesis that cellular competition and chemotherapy resistance work as dynamic evolutionary forces that shape the clonal architecture of ALL^{8–10}. To test this hypothesis we generated a knock-in mouse model (*Nt5c2*^{+/co-R367Q}) for conditional expression of *Nt5c2*(R367Q) (Extended Data Fig. 2), the most common *NT5C2* mutation found in relapsed ALL^{2,3}, and generated primary NOTCH1-induced *Rosa26*^{+/CreERT2}*Nt5c2*^{+/co-R367Q} T-ALL tumours^{17,18} (Extended Data Fig. 2) with conditional tamoxifen-inducible expression of *Nt5c2*(R367Q) (Fig. 1a and Extended Data Fig. 2). Treatment of isogenic *Nt5c2* wild-type (*Nt5c2*^{+/co-R367Q}, vehicle-treated) and *Nt5c2*^{+/R367Q} mutant (*Nt5c2*^{+/co-R367Q}, 4-hydroxy-tamoxifen-treated) leukaemia cells with increasing concentrations of 6-MP showed overt resistance to thiopurine chemotherapy specifically in *Nt5c2*^{+/R367Q} mutant cells (Fig. 1b). Moreover, *Nt5c2*^{+/R367Q} mutant cells were positively selected in a dose-dependent manner over isogenic wild-type *Nt5c2*^{+/co-R367Q} tumour cells under 6-MP treatment *in vitro* (Fig. 1c). Treatment of mice harbouring isogenic *Nt5c2*^{+/co-R367Q} (vehicle treated, wild-type group) or *Nt5c2*^{+/R367Q} (tamoxifen treated, mutant group) leukaemias with 6-MP produced a dose-dependent response in *Nt5c2*^{+/co-R367Q} wild-type tumours and overt resistance with progression in *Nt5c2*^{+/R367Q} mutant leukaemias (Fig. 1d, e and Extended Data Fig. 2g). Moreover, 6-MP treatment of mixed tumour populations of isogenic wild-type *Nt5c2*^{+/co-R367Q} and mutant *Nt5c2*^{+/R367Q} lymphoblasts demonstrated positive selection *in vivo* of cells harbouring the *Nt5c2*(R367Q)-encoding mutant allele (Extended Data Fig. 2h). These results support a direct role for *NT5C2*(R367Q) as a driver of 6-MP resistance *in vivo* and are concordant with the strong association of *NT5C2* mutations with early relapse and progression during 6-MP maintenance therapy in the clinic^{2,3}.

Recent genomic studies of matched diagnostic and relapsed ALL samples support the hypothesis that relapsed leukaemia emerges from the expansion of pre-existing resistant populations present as minor

¹Institute for Cancer Genetics, Columbia University, New York, New York 10032, USA. ²Department of Systems Biology, Columbia University, New York, New York 10032, USA. ³Rutgers Cancer Institute, Rutgers University, New Brunswick, New Jersey 08903, USA. ⁴Princess Maxima Center for Pediatric Oncology, Utrecht, 3584 CT, the Netherlands. ⁵Department of Human Genetics, Radboud University Medical Center and Radboud Institute for Molecular Life Sciences, Nijmegen, 6525 GA, the Netherlands. ⁶Department of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ⁷Department of Pediatrics, Columbia University Medical Center, New York, New York 10032, USA. ⁸Department of Hematology-Oncology, Saitama Children's Medical Center, Saitama 339-8551, Japan. ⁹Onco-Hematology Division, Department, Salute della Donna e del Bambino (SDB), University of Padua, 35128 Padua, Italy. ¹⁰Department of Pathology and Laboratory Medicine, Nationwide Children's Hospital, Columbus, Ohio 43205, USA. ¹¹Department of Pathology, Ohio State University School of Medicine, Columbus, Ohio 43210, USA. ¹²Department of Pediatrics, Ohio State University School of Medicine, Columbus, Ohio 43210, USA. ¹³Children's Oncology Group, Arcadia, California 91006, USA. ¹⁴Department of Pediatrics, University of California, San Francisco, California 94143, USA. ¹⁵Helen Diller Family Comprehensive Cancer Center, San Francisco, California 94115, USA. ¹⁶Department of Pediatric Oncology/Hematology, Charité-Universitätsmedizin Berlin, Berlin, 10117, Germany. ¹⁷Department of Biomedical Informatics, Columbia University, New York, New York 10032, USA. ¹⁸Department of Pathology and Cell Biology, Columbia University Medical Center, New York, New York 10032, USA. †Present addresses: Regeneron Pharmaceuticals, Tarrytown, New York, New York 10591, USA (G.T.); PsychoGenics, Paramus, New Jersey 07652, USA (A.A.-I.).

*These authors contributed equally to this work.

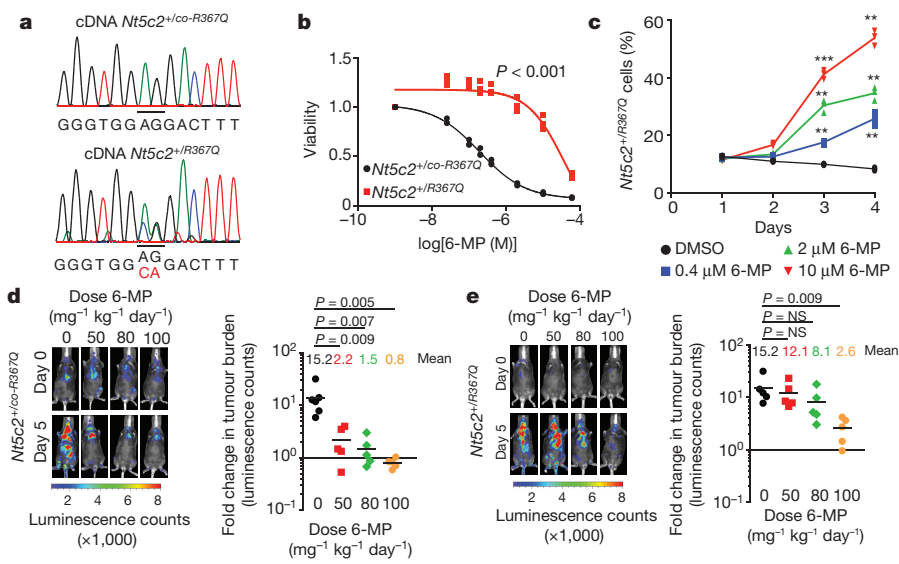


Figure 1 | Expression of *Nt5c2*(R367Q) in a NOTCH1-induced mouse model of ALL induces resistance to 6-MP. **a**, cDNA sequencing chromatograms of *Nt5c2*^{+/co-R367Q} wild-type and *Nt5c2*^{+/R367Q} isogenic T-ALL cells. Data are representative of results from more than two experiments. **b**, Cell viability of isogenic *Nt5c2*^{+/co-R367Q} and *Nt5c2*^{+/R367Q} T-ALL cells treated with increasing concentrations of 6-MP ($n = 3$ biological replicates). **c**, Change in the percentage of *Nt5c2*^{+/R367Q} T-ALL cells over time in a mixed culture with isogenic *Nt5c2*^{+/co-R367Q} cells treated with 6-MP ($n = 3$ biological replicates). DMSO, dimethylsulfoxide. **d**, **e**, Tumour burden in mice allografted with *Nt5c2*^{+/co-R367Q} (d) or isogenic *Nt5c2*^{+/R367Q} (e) leukaemia cells and treated with 6-MP. Data for the vehicle groups are from 6 (d) or 5 (e) mice and from 5 mice for the treatment groups. $**P \leq 0.01$, $***P \leq 0.001$, two-tailed Student's *t*-test. NS, not significant.

subclones at the time of diagnosis¹⁹. To evaluate further the role of NT5C2 as a driver of clonal progression and relapse in ALL, we used ultra-deep sequencing with unique-molecular-identifier barcoding (4,100 \times coverage) to analyse the presence of NT5C2 mutations in 14 diagnostic DNA samples from cases showing acquired NT5C2 mutations at relapse. Notably, these analyses (1:1,000 sensitivity) failed to detect the corresponding relapse-associated NT5C2 mutant allele at the time of diagnosis (Extended Data Table 1). NT5C2(R367Q) allele-specific quantitative PCR (qPCR) ($n = 9$) (1:1,000 sensitivity) yielded similar negative results (Extended Data Table 1). Moreover, in one case bearing the NT5C2(R39Q) mutation at the time of relapse, droplet PCR analysis (1:20,000 sensitivity) detected the presence of this mutation during complete remission 37 days prior to the emergence of clinical relapse (Extended Data Table 1). Before then, and at the time of diagnosis, the signal for this mutation (0.00064%) was below the established sensitivity of the assay (0.005%). In a separate case we detected a NT5C2(P414A) mutation in first relapse and a second NT5C2(R39Q) variant in second relapse. In this patient, the NT5C2(P414A) mutation was not detectable by droplet PCR analysis at the time of diagnosis, whereas the mutant allele encoding NT5C2(R39Q) was detected below the 0.005% detection threshold at 0.0024–0.0031% frequency. However, analysis of bone marrow at the time of first relapse detected a NT5C2(R39Q) subclonal population (0.0058%) in addition to the NT5C2(P414A) clone. These NT5C2(R39Q) mutant cells expanded (0.0224%) in a serial sample obtained during a second complete remission 60 days later, while the NT5C2(P414A) mutant clone decreased, becoming clonal at the time of second relapse 50 days later (Extended Data Table 1). These results suggest that NT5C2 mutations can be detected in complete-remission samples before relapse, yet, if present in the clonal repertoire at the time of diagnosis, they represent quantitatively minor populations below the sensitivity of molecular assays.

Resistance-driving mutations have been linked to enhanced leukaemia growth and proliferation, clonal expansion at early stages of tumour development and increased leukaemia stem-cell activity^{20–22}. However, studies of resistance to bacterial antibiotics have uncovered frequent examples of evolutionary trade-offs in which the acquisition of drug resistance is coupled with a reduced-fitness phenotype²³. In this context, we noted that in the absence of chemotherapy, *Nt5c2*^{+/R367Q} mouse tumour cells showed decreased proliferation *in vitro*, a delayed entry into the S phase of the cell cycle (Fig. 2a, b) and delayed tumour progression *in vivo* compared with wild-type *Nt5c2*^{+/co-R367Q} isogenic controls (Fig. 2c). Moreover, limiting dilution transplantation assays demonstrated a 17-fold reduction of leukaemia-initiating cell activity in mutant *Nt5c2*^{+/R367Q} tumour cells (Fig. 2d–e and Extended Data Table 2). Of note, allele expression analysis of tumours recovered from

mice transplanted with *Nt5c2*^{+/R367Q} leukaemia lymphoblasts showed decreased expression of the mutant *Nt5c2* transcripts, suggesting downregulation of the mutant allele encoding Nt5c2(R367Q) during tumour progression in the absence of 6-MP (Extended Data Fig. 3). These results support the hypothesis that Nt5c2(R367Q) imposes a notable fitness cost to leukaemia lymphoblasts.

Given the role of NT5C2 in the degradation and export of purine nucleotides¹⁵, we examined whether imbalances in the intracellular purine-nucleotide pool could mediate the loss-of-fitness phenotype observed in *Nt5c2*^{+/R367Q} mutant leukaemia cells. Broad-based metabolomic analysis showed that NT5C2 activation in *Nt5c2*^{+/R367Q} ALL lymphoblasts leads to decreased intracellular levels of NT5C2 substrates (inosine monophosphate, xanthine monophosphate and guanosine monophosphate) and accumulation of downstream nucleotide products and their metabolites (inosine, hypoxanthine, xanthosine, xanthine, guanine and uric acid) in conditioned media (Fig. 3 and Supplementary Tables 1, 2). Similarly, expression of NT5C2(R367Q)

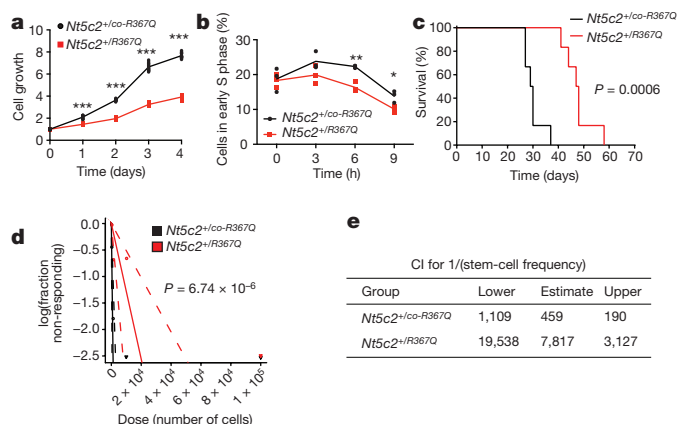


Figure 2 | *Nt5c2*(R367Q) expression impairs proliferation and leukaemia-initiating cell activity in ALL. **a**, *In vitro* growth (fold change) of isogenic *Nt5c2*^{+/co-R367Q} wild-type and *Nt5c2*^{+/R367Q} mutant mouse T-ALL cells. **b**, Cell cycle progression of *Nt5c2*^{+/co-R367Q} and *Nt5c2*^{+/R367Q} mouse T-ALL cells. **c**, Kaplan–Meier survival curve of mice harbouring *Nt5c2*^{+/co-R367Q} and *Nt5c2*^{+/R367Q} isogenic leukaemias ($n = 6$ per group). **d**, Leukaemia-initiating cell analysis in mice bearing *Nt5c2*^{+/co-R367Q} or isogenic *Nt5c2*^{+/R367Q} leukaemia cells ($n = 6$ mice per group). **e**, Confidence intervals (CI) showing 1/(stem-cell frequency) based on d. **a**, **b**, Data are from three biological replicates. $*P \leq 0.05$, $**P \leq 0.005$, $***P \leq 0.001$, two-tailed Student's *t*-test (**a**, **b**) or two-sided log-rank test (**c**).

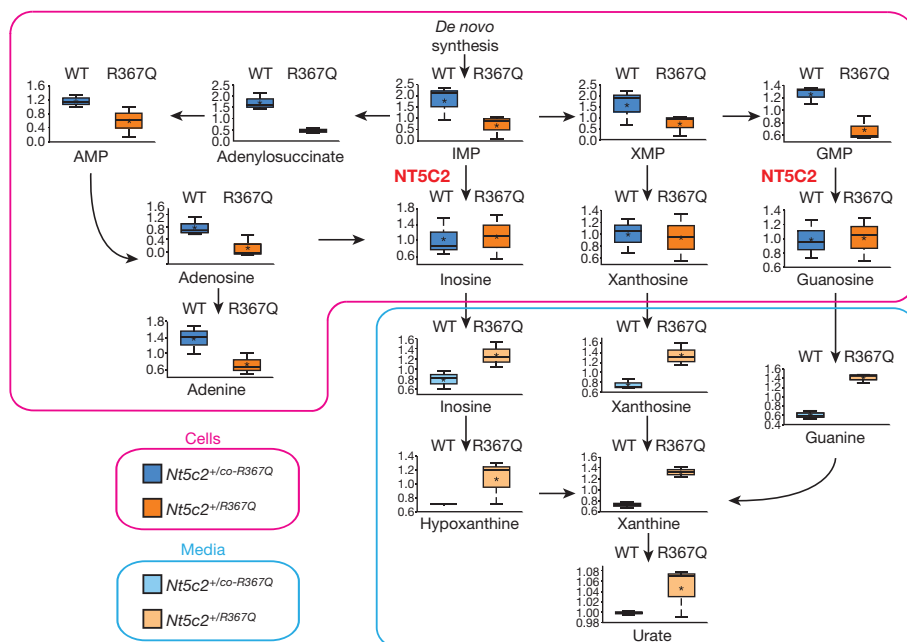


Figure 3 | *Nt5c2*(R367Q) decreases the intracellular purine nucleoside pool and increases secretion of purines in ALL cells. Diagram of the purine *de novo* biosynthesis and salvage pathways, showing gas chromatography–mass spectrometry and liquid chromatography–tandem mass spectrometry metabolic profiles (mass spectrometry scaled intensity, arbitrary units) of wild-type *Nt5c2*^{+/co-R367Q} (WT) and *Nt5c2*^{+/R367Q} mutant (R367Q) isogenic primary mouse T-ALL cells and their corresponding conditioned media ($n = 3$ biological replicates). Box plots represent the upper quartile to lower quartile distribution. Asterisks indicate median values, horizontal lines indicate the maximum and minimum values.

in human T-ALL (CUTLL1) and B precursor ALL (REH) cell lines resulted in depletion of intracellular purine nucleotides and increased levels of purine metabolites in the culture media (Extended Data Fig. 4 and Supplementary Tables 3, 4). Increased extracellular purine metabolites are consistent with the described activity of NT5C2 in promoting the export of purine nucleotides¹⁵ and might result in potential non-cell autonomous satellite effects modulating nucleotide metabolism and the response to 6-MP in by-standing wild-type NT5C2 cells.

A corollary of these findings is that because of this metabolic imbalance, gain-of-function NT5C2 mutations could be negatively selected during ALL tumour initiation and early disease progression, a time when clonal evolution is driven primarily by competition for microenvironment resources with normal haematopoietic stem and progenitor cells first, and then between different leukaemia clones²⁴. Consistent with this model, integrated sequential network (ISN)²⁵ analysis of mutation dynamics from diagnostic and relapse mutation data

identified NT5C2 mutations as late events in the clonal evolution of ALL (Extended Data Fig. 5).

We hypothesized that gain-of-function relapse-associated NT5C2 mutations could result in increased leukaemia dependence on purine synthesis, rendering leukaemia lymphoblasts more sensitive to drugs targeting this pathway. Indeed, acquired drug resistance in bacteria can be accompanied by collateral sensitivity to an alternative antibiotic agent²³. To test this possibility, we analysed the response of *Nt5c2*^{+/R367Q} mutant ALL lymphoblasts to mizoribine, an inhibitor of inosine-5'-monophosphate dehydrogenase (IMPDH), a rate-limiting enzyme required for the synthesis of guanine nucleotides²⁶. Notably, these experiments demonstrated significantly increased sensitivity to mizoribine in *Nt5c2*^{+/R367Q} mutant leukaemia cells *in vitro* compared to *Nt5c2*^{+/co-R367Q} wild-type isogenic controls (Fig. 4a and Extended Data Fig. 6). Moreover, guanosine supplementation in the media rescued the effects of mizoribine in

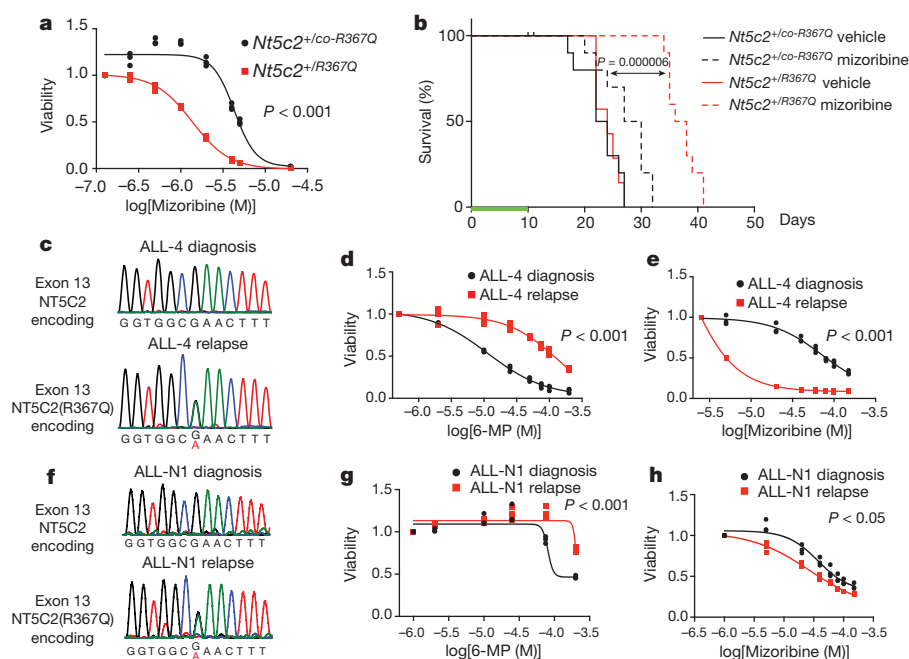


Figure 4 | Collateral sensitivity to IMPDH inhibition in NT5C2(R367Q) mutant tumour cells. a, Cell viability of isogenic *Nt5c2*^{+/co-R367Q} and *Nt5c2*^{+/R367Q} T-ALLs treated with mizoribine. b, Kaplan–Meier survival curve of mice harbouring *Nt5c2*^{+/co-R367Q} and *Nt5c2*^{+/R367Q} isogenic leukaemias ($n = 10$ mice per group) after initial treatment with mizoribine or vehicle (green bar). c, f, DNA sequencing chromatograms corresponding to two matched primary human T-ALL xenografts obtained from samples taken at the time of diagnosis (wild-type NT5C2) and after relapse (NT5C2(R367Q)) (matched pairs are labelled as ALL-4 (c–e) and ALL-N1 (f–h)). d, g, Cell viability of samples evaluated in c and f showing resistance to 6-MP in relapsed T-ALL xenograft cells. e, h, Cell viability in samples analysed in c, f, d and g showing collateral sensitivity to mizoribine in relapsed T-ALL xenograft cells. Data in a, d, e, g and h are from three biological replicates and P values were calculated using two-tailed Student's t -test. P value in b was calculated with a two-sided log-rank test.

Nt5c2^{+/co-R367Q} wild-type and *Nt5c2*^{+/R367Q} mutant lymphoblasts, supporting a mechanistic role for nucleotide depletion in the activity of this drug and its synthetic lethal interaction with the mutant allele encoding *Nt5c2*(R367Q) (Extended Data Fig. 6). Similar differential responses to mizoribine in wild-type *Nt5c2*^{+/co-R367Q} cells and in *Nt5c2*^{+/R367Q} mutant cells were observed *in vivo* in a subcutaneous lymphoma model (Extended Data Fig. 6). Furthermore, treatment of *Nt5c2*^{+/R367Q} leukaemia-bearing mice with mizoribine induced a marked *in vivo* anti-leukaemic response, with significantly improved survival compared with isogenic *Nt5c2*^{+/co-R367Q} wild-type controls ($P < 0.0001$) (Fig. 4b). Similarly, expression of gain-of-function *NT5C2* mutations (R238W, K359Q, R367Q and D407A) associated with relapse in CUTLL1 and REH ALL cells induced resistance to 6-MP and increased their sensitivity to mizoribine (Extended Data Figs 7, 8). As before, guanosine supplementation ameliorated the anti-leukaemic effects of mizoribine in this experiment, providing evidence that depletion of nucleotides is the mechanism of action for this drug (Extended Data Figs 7, 8). Additionally, knockdown of *IMPDH2*, the gene encoding the main IMPDH isoform expressed in proliferating tissues and tumour cells, led to decreased growth in *NT5C2*(R367Q)-expressing REH and CUTLL1 cells compared to wild-type *NT5C2*-expressing lymphoblasts (Extended Data Figs 7, 8). We also observed resistance to 6-MP with increased sensitivity to mizoribine in leukaemia cells from two human primary xenografts harbouring the mutant *NT5C2*(R367Q)-encoding allele compared to matched wild-type *NT5C2* ALL blasts derived from samples obtained at the time of diagnosis (Fig. 4c–h). Moreover, immunodeficient mice transplanted with an *NT5C2*(R367Q) xenograft derived from a relapsed patient showed decreased tumour burden and tumour infiltration following mizoribine treatment compared to mice transplanted with matched wild-type *NT5C2* xenograft cells derived from samples taken at the time of diagnosis (Extended Data Fig. 6).

These results document fitness cost and acquired resistance to 6-MP as evolutionary forces that drive the clonal evolution dynamics and selection of relapse-associated *NT5C2* mutations in ALL, highlight the relevance of nucleotide export in the control of nucleotide homeostasis²⁷ and in the context of antimetabolite therapy²⁸, and identify collateral sensitivity to IMPDH inhibition as a potentially relevant vulnerability in *NT5C2*-mutant leukaemia.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 December 2016; accepted 30 November 2017.

Published online 17 January 2018.

- Hunger, S. P. & Mullighan, C. G. Acute lymphoblastic leukemia in children. *N. Engl. J. Med.* **373**, 1541–1552 (2015).
- Tzoneva, G. *et al.* Activating mutations in the *NT5C2* nucleotidase gene drive chemotherapy resistance in relapsed ALL. *Nat. Med.* **19**, 368–371 (2013).
- Meyer, J. A. *et al.* Relapse-specific mutations in *NT5C2* in childhood acute lymphoblastic leukemia. *Nat. Genet.* **45**, 290–294 (2013).
- Gerby, B. *et al.* Expression of CD34 and CD7 on human T-cell acute lymphoblastic leukemia discriminates functionally heterogeneous cell populations. *Leukemia* **25**, 1249–1258 (2011).
- Cox, C. V., Diamanti, P., Evelyn, R. S., Kearns, P. R. & Blair, A. Expression of CD133 on leukemia-initiating cells in childhood ALL. *Blood* **113**, 3287–3296 (2009).
- Konopleva, M. *et al.* Stromal cells prevent apoptosis of AML cells by up-regulation of anti-apoptotic proteins. *Leukemia* **16**, 1713–1724 (2002).
- Hawkins, E. D. *et al.* T-cell acute leukaemia exhibits dynamic interactions with bone marrow microenvironments. *Nature* **538**, 518–522 (2016).
- Mullighan, C. G. *et al.* Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* **322**, 1377–1380 (2008).
- Ma, X. *et al.* Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. *Nat. Commun.* **6**, 6604 (2015).
- Oshima, K. *et al.* Mutational landscape, clonal evolution patterns, and role of RAS mutations in relapsed acute lymphoblastic leukemia. *Proc. Natl Acad. Sci. USA* **113**, 11306–11311 (2016).
- Li, B. *et al.* Negative feedback-defective PRPS1 mutants drive thiopurine resistance in relapsed childhood ALL. *Nat. Med.* **21**, 563–571 (2015).

- Mullighan, C. G. *et al.* *CREBBP* mutations in relapsed acute lymphoblastic leukaemia. *Nature* **471**, 235–239 (2011).
- Malinowska-Ozdowy, K. *et al.* *KRAS* and *CREBBP* mutations: a relapse-linked malicious liaison in childhood high hyperdiploid acute lymphoblastic leukemia. *Leukemia* **29**, 1656–1667 (2015).
- Spychala, J., Madrid-Marina, V. & Fox, I. H. High K_m soluble 5'-nucleotidase from human placenta. Properties and allosteric regulation by IMP and ATP. *J. Biol. Chem.* **263**, 18759–18765 (1988).
- Gazziola, C., Ferraro, P., Moras, M., Reichard, P. & Bianchi, V. Cytosolic high K_m 5'-nucleotidase and 5'(3')-deoxyribonucleotidase in substrate cycles involved in nucleotide metabolism. *J. Biol. Chem.* **276**, 6185–6190 (2001).
- Brouwer, C. *et al.* Role of 5'-nucleotidase in thiopurine metabolism: enzyme kinetic profile and association with thio-GMP levels in patients with acute lymphoblastic leukemia during 6-mercaptopurine treatment. *Clin. Chim. Acta* **361**, 95–103 (2005).
- Schroeter, E. H., Kisslinger, J. A. & Kopan, R. Notch-1 signalling requires ligand-induced proteolytic release of intracellular domain. *Nature* **393**, 382–386 (1998).
- Herranz, D. *et al.* A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat. Med.* **20**, 1130–1137 (2014).
- Ferrando, A. A. & López-Otín, C. Clonal evolution in leukemia. *Nat. Med.* **23**, 1135–1145 (2017).
- Clappier, E. *et al.* Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. *J. Exp. Med.* **208**, 653–661 (2011).
- Wong, T. N. *et al.* Rapid expansion of preexisting nonleukemic hematopoietic clones frequently follows induction therapy for *de novo* AML. *Blood* **127**, 893–897 (2016).
- Shlush, L. I. *et al.* Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature* **547**, 104–108 (2017).
- Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat. Rev. Microbiol.* **8**, 260–271 (2010).
- Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
- Wang, J. *et al.* Tumor evolutionary directed graphs and the history of chronic lymphocytic leukemia. *eLife* **3**, (2014).
- Gan, L. *et al.* The immunosuppressive agent mizoribine monophosphate forms a transition state analogue complex with inosine monophosphate dehydrogenase. *Biochemistry* **42**, 857–863 (2003).
- Reaves, M. L., Young, B. D., Hosios, A. M., Xu, Y. F. & Rabinowitz, J. D. Pyrimidine homeostasis is accomplished by directed overflow metabolism. *Nature* **500**, 237–241 (2013).
- Ser, Z. *et al.* Targeting one carbon metabolism with an antimetabolite disrupts pyrimidine homeostasis and induces nucleotide overflow. *Cell Reports* **15**, 2367–2376 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to R. Kopan for the ΔE -NOTCH1 construct and T. Ludwig for the *ROSA26*^{Cre-ERT2/+} mouse. This work was supported by the Leukemia & Lymphoma Society Quest for Cures (R0749-14) and Translational Research (6455-15; 6531-18) awards (A.A.F.), an Innovative Research Award from the Alex Lemonade Stand Foundation (A.A.F.), the Chemotherapy Foundation (A.A.F.), National Institutes of Health (NIH) grants R35 CA210065 (A.A.F.), R01 CA206501 (A.A.F.), U54 CA193313 (R.R.), R01 CA185486 (R.R.), U54 CA209997 (R.R.), U10 CA98543 (J.M.G., M.L.L.), P30 CA013696, the Human Specimen Banking Grant U24 CA114766 (J.M.G.), the Stewart Foundation (R.R.) and the American Lebanese Syrian Association Charities of St Jude Children's Research Hospital. G.T. was supported by a HHMI International Student Research Fellowship. M.S.M. was supported by a Rally Foundation fellowship. C.L.D. was supported by NIH/NCI T32-CA09503. J.Y. was supported by the China Scholarship Council (CSC 201304910347) and the Ter Meulen Grant of the Royal Netherlands Academy of Arts and Sciences. E.W. was supported by the Dutch Cancer Society (KUN2012-5366).

Author Contributions G.T. and C.L.D. performed biochemical, cellular and animal studies. M.S.-M. and K.O. helped in experimental therapeutic experiments. A.A.-I. and H.K. analysed deep sequencing data. C.J.M. performed ISN analysis. M.L.S., M.K., K.K., M.P., G.B., J.M.G.-F. and M.L.L. provided clinical specimens. J.Y., E.W. and I.I. performed and analysed droplet PCR analyses. R.K.-S. provided clinical samples and correlative analyses of clinical data. C.G.M. supervised droplet PCR analyses; R.R. supervised deep sequencing and ISN analyses. A.A.F. designed the study, supervised the research and wrote the manuscript with G.T. and C.L.D.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.A.F. (af2196@columbia.edu).

METHODS

Patient samples. DNA samples from leukaemic ALL blasts obtained at diagnosis and after relapse and matched remission lymphocytes were provided by the Hemato-Oncology Laboratory at University of Padua, Italy; the Children's Oncology Group in the Department of Hematology/Oncology at Saitama Children's Medical Center, Saitama, Japan and St Jude Children's Research Hospital. Informed consent was obtained at study entry and samples were collected under the supervision of local Institutional Review Boards for participating institutions and analysed under the supervision of the Columbia University Medical Center Institutional Review Board (Protocol Number: IRB-AAAB3250). Research was conducted in compliance with ethical regulations.

Cell lines and cell culture procedures. We performed cell culture in a humidified atmosphere at 37°C under 5% CO₂. We harvested primary mouse tumour cells from the spleens of leukaemic mice by processing spleens through a 70-µm mesh to obtain single-cell suspensions and incubated cells with red blood cell lysis buffer. Tumour cells were then placed in culture in Opti-MEM media supplemented with 10% fetal bovine serum (FBS), 100 U ml⁻¹ penicillin G, 100 µg ml⁻¹ streptomycin, 55 µM β-mercaptoethanol, 10 ng ml⁻¹ mouse IL-7 and 10 ng ml⁻¹ human IL-2. Subsequent passages of tumour cells did not include IL-2. We passaged and harvested primary human xenograft T-ALL cells from the spleens of NRG (NOD.Cg-Rag1^{tm1Mom}/J2rg^{tm1Wjl}/SzJ, Jackson Laboratory) mice and cultured them in RPMI media supplemented with 20% FBS, 100 U ml⁻¹ penicillin G, 100 µg ml⁻¹ streptomycin and 10 ng ml⁻¹ human IL-7. We purchased HEK293T cells for viral production and REH cells from American Type Culture Collection. The CUTLL1 cell line, which was generated by continuous culture of T-cell lymphoblastic pleural effusion cells from a patient in relapse, has been characterized and reported previously²⁹. We grew HEK293T cells in DMEM media supplemented with 10% FBS, 100 U ml⁻¹ penicillin G and 100 µg ml⁻¹ streptomycin for up to two weeks. We cultured CUTLL1 and REH cells in RPMI-1640 media supplemented with 10% FBS, 100 U ml⁻¹ penicillin G and 100 µg ml⁻¹ streptomycin. Cell lines were regularly authenticated and tested for mycoplasma contamination.

Drugs. We purchased tamoxifen, guanosine, 4-hydroxytamoxifen, 6-mercaptopurine (6-MP) and mizoribine from Sigma-Aldrich. For *in vitro* assays we dissolved 4-hydroxytamoxifen in 100% ethanol, guanosine in DMSO, 6-MP in DMSO and mizoribine in PBS. For *in vivo* studies we resuspended 100 mg tamoxifen in 100 µl of ethanol and added corn oil to reach a final concentration of 3 mg 100 µl⁻¹. We then rotated the tamoxifen suspension for 1 h at 55°C and froze it in aliquots at -20°C. We administered tamoxifen as a single 100 µl intraperitoneal injection per mouse. For intraperitoneal injections of 6-MP we prepared frozen aliquots of 5 mg ml⁻¹ 6-MP in 0.1 M NaOH and immediately before each round of treatment we prepared fresh final solutions of 6-MP by buffering the stock solution down to pH 8 with 0.2 M NaH₂PO₄. This resulted in a 6-MP concentration of 3.48 mg ml⁻¹, which we diluted to various final concentrations using a solution made from 0.05 M NaOH and 0.2 M NaH₂PO₄ adjusted to pH 8. We administered 6-MP as 25 mg kg⁻¹, 40 mg kg⁻¹ and 50 mg kg⁻¹ all twice a day. We prepared vehicle by dissolving 0.254 g NaCl in 50 ml 0.05 M NaOH and adjusting the pH to 8 with 0.2 M NaH₂PO₄. For intraperitoneal injections we dissolved mizoribine (TCI America and Toronto Research Chemicals) in PBS at 10 mg ml⁻¹ or 15 mg ml⁻¹ and froze aliquots to be thawed before treatment. We adjusted injection volume to correct for any differences in weight between individual mice.

Plasmid and vectors. We obtained MigR1_ΔE-NOTCH1_GFP from R. Kopan, sh-TURBOGFP and pLKO.1_IMPDI2_shRNA (clone ID: NM_000884.1-360s1c1) from Sigma Aldrich's Mission shRNA library and FUW-mCherry-Puro-Luc from ref. 30. We generated the NT5C2 R238W, K359Q, R367Q and D407A mutations in the pLOC-NT5C2 plasmid² by site-directed mutagenesis using the QuikChange II XL Site-Directed Mutagenesis kit (Stratagene) according to the manufacturer's instructions.

Retroviral and lentiviral infections. We transfected retroviral or lentiviral plasmids together with gag-pol (pCMV ΔR8.91) and V-SVG (pMD.G VSVG) expressing vectors into HEK293T cells using JetPEI transfection reagent (Polyplus). We collected viral supernatants after 48 h and used them to infect mouse bone marrow progenitors, human cell lines, or primary tumour cells by spinoculation with 4 µg ml⁻¹ Polybrene Infection/Transfection Reagent (Fisher Scientific). We selected infected primary mouse tumour cells or human cell lines with 1 mg ml⁻¹ blasticidin (InvivoGen) for 14 days.

Mice and animal procedures. We maintained all animals in specific pathogen-free facilities at the Irving Cancer Research Center at Columbia University Medical Campus. The Columbia University Institutional Animal Care and Use Committee (IACUC) approved all animal procedures. Animal experiments were conducted in compliance with all relevant ethical regulations. Animals were euthanized upon showing symptoms of clinically overt disease (do not feed, lack of activity, abnormal grooming behaviour, hunch back posture) or excessive weight loss (10–15% body weight loss over a week) and before reaching the maximum permitted tumour

burden of 90% blasts in the bone marrow. To generate conditional Nt5c2(R367Q) knock-in mice we used homologous recombination in C57BL/6 embryonic stem cells to introduce a point mutation (AG→CA) in exon 14 that caused the R367Q substitution (two nucleotide changes were introduced to replace the mouse R367 codon (AGA) with a glutamine-coding codon (CAA)) as well as a loxP-flanked wild-type mini-gene cassette (1958 bp, inserted 233 bp upstream of exon 14) composed of the fusion of exons 14–18 and flanking genomic sequences upstream of exon 14 and downstream of exon 18. Immediately downstream of the mini-gene we introduced a FRT-flanked neomycin selection cassette. We generated chimaeras in C57BL/6 albino blastocysts using three independent knock-in embryonic stem cell clones identified by PCR analysis and verified by Southern blot. We verified germ-line transmission in the offspring of highly chimaeric male mice crossed with C57BL/6 females. To remove the neomycin selection cassette we crossed mice harbouring the targeting construct with a Flp germ line deleter line (B6;SJL-Tg(ACTFLPe)9205Dym/J, Jackson Laboratory) and crossed the resulting mice with wild-type C57BL/6 to breed out the Flp allele. To generate inducible knock-in mice we bred animals harbouring the Nt5c2^{co-R367Q} allele with Rosa26^{+/creERT2} mice, which express a tamoxifen-inducible form of the Cre recombinase from the ubiquitous Rosa26 locus³¹.

To generate NOTCH1-induced T-ALL tumours in mice, we performed retroviral transduction of bone marrow cells (from Rosa26^{+/creERT2}Nt5c2^{+/co-R367Q} mice) enriched in lineage negative cells isolated using magnetic beads (Lineage Cell Depletion Kit, Miltenyi Biotec) with retroviruses expressing an activated form of the NOTCH1 oncogene (ΔE-NOTCH1)¹⁷ and the green fluorescent protein (GFP) and transplanted them via intravenous injection into lethally irradiated isogenic recipients (6–8-week-old C57BL/6 females, Taconic Farms) as previously described^{18,32}.

We assessed T-ALL tumour development by monitoring CD4⁺CD8⁺GFP⁺ cells in peripheral blood by flow cytometry. In brief, we incubated blood samples with red blood cell lysis buffer (155 mM NH₄Cl, 10 mM KHCO₃, 0.1 mM EDTA) for 5 min at room temperature three times before staining with APC-Cy7-conjugated antibodies against mouse CD4 (BD Pharmingen-552051) and PE-Cy7-conjugated antibodies against mouse CD8a (eBioscience-25-0081-82). Flow cytometry analyses were performed in a FACSCanto flow cytometer (BD Biosciences) and analysed with FlowJo software (FlowJo LLC).

For all subsequent *in vivo* studies, we harvested fresh Rosa26^{+/creERT2}Nt5c2^{+/co-R367Q} T-ALL tumour cells from the spleens of donor mice and transplanted them into sublethally irradiated (4 Gy) secondary recipients (C57BL/6 females, 6–8 weeks old, Taconic Farms). Animals were randomly assigned to different treatment groups and no blinding was done. For survival and leukaemia-initiating cell experiments, we treated mice with tamoxifen (3 mg via intraperitoneal injection) two days after transplantation to induce mini-gene-cassette deletion and expression of the Nt5c2 allele encoding the R367Q mutation in the leukaemic cells, or with corn oil vehicle in the control group (*n* = 6 mice per group). Mice were then observed for incidence and time of onset of leukaemia.

To detect tamoxifen-inducible mini-gene deletion, we purified DNA from primary Rosa26^{+/creERT2}Nt5c2^{+/co-R367Q} mouse tumour cells treated with 1 µM 4-hydroxytamoxifen or ethanol vehicle (*in vitro* experiments) and tamoxifen or corn oil vehicle (*in vivo* experiments) and then performed PCR-amplification with a three-primer reaction: (i) the minigene cassette (primer immediately upstream of proximal loxP site and reverse primer in exon 17) and (ii) the deleted mini-gene and wild-type alleles (primer immediately upstream of proximal loxP site and reverse primer in intron 14). The deleted and wild-type alleles differ by the size of the remaining loxP site (49 bp). We visualized PCR products resolved by electrophoresis in a 1.5% agarose gel with ethidium bromide.

To detect tamoxifen-inducible expression of mRNA corresponding to the Nt5c2 allele encoding the R367Q mutation, we purified total RNA from mouse tumour cells with the RNeasy kit (Qiagen), prepared complementary DNA (cDNA) by reverse transcription using the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen) and PCR amplified the Nt5c2 exon 14 cDNA region using primers spanning neighbouring exons following standard procedures. We analysed the resulting PCR products by dideoxy DNA sequencing to verify the expression of the engineered nucleotide substitutions in the Nt5c2 allele encoding the R367Q mutation.

For experimental therapeutics treatment studies, we used Rosa26^{+/creERT2}Nt5c2^{+/co-R367Q} T-ALL tumour cells infected with lentiviral particles expressing the red cherry fluorescent protein and luciferase (FUW-mCherry-Luc-puro). We transplanted luciferase-expressing Rosa26^{+/creERT2}Nt5c2^{+/co-R367Q} T-ALL tumour cells into C57BL/6 recipients by intravenous injection and monitored tumour development by *in vivo* luminescence bioimaging with the *In vivo* Imaging System (IVIS, Xenogen) and by flow cytometry using analysis of GFP⁺ cells in peripheral blood. Once mice had 50% GFP positive cells in the peripheral blood and a detectable baseline tumour burden by bioluminescence, we treated them with tamoxifen

(3 mg, intraperitoneal injection) or corn oil vehicle as described above. Two days later we initiated treatment with a range of doses of 6-MP (0, 50, 80, 100 mg kg⁻¹ per day) via intraperitoneal injection for five consecutive days ($n = 5$ mice per group). We monitored disease progression and response to chemotherapy by bioluminescence imaging on days 0, 3 and 6 after the start of 6-MP treatment. We euthanized mice on day 6 and analysed GFP⁺ tumour infiltration in the spleen by flow cytometry. To assess mizoribine response *in vivo* we treated *Rosa26⁺/creERT2* *Nt5c2^{+/co-R367Q}* leukaemia bearing mice 48 h following tamoxifen or corn oil vehicle treatment (as described above) with 40 mg kg⁻¹ mizoribine or PBS vehicle ($n = 10$ per group) via intraperitoneal injection for ten consecutive days. Mice were then observed for incidence and time of onset of leukaemia.

For experimental therapeutic treatment studies in a subcutaneous setting, *Rosa26⁺/creERT2* *Nt5c2^{+/co-R367Q}* T-ALL tumour cells infected with lentiviral particles expressing the red cherry fluorescent protein and luciferase (FUW-mCherry-Luc-puro) were treated with 1 μ M 4-hydroxytamoxifen or ethanol vehicle *in vitro*, mixed with an equal volume of Corning Matrigel Membrane Matrix (Fisher Scientific) and injected (10⁶ cells) into the flanks of female C57BL/6 mice. Upon detectable baseline tumour burden by bioluminescence, mice were treated intraperitoneally with PBS vehicle or mizoribine (20, 40, 75 or 100 mg kg⁻¹ per day, $n = 4$ per dose) for 5 consecutive days. We monitored tumour progression and response to mizoribine by bioluminescence imaging on days 0 and 6 after the start of mizoribine treatment. Subcutaneous tumours were not allowed to exceed 20 mm in diameter.

To evaluate the competitive selection of *Nt5c2^{+/R367Q}* cells *in vivo* we mixed wild-type *Nt5c2^{+/co-R367Q}* and *Nt5c2^{+/R367Q}* mutant mouse tumour cells at a 1:10, 1:100 and 1:1,000 *Nt5c2^{+/R367Q}:Nt5c2^{+/co-R367Q}* ratios and transplanted into C57BL/6 recipients by intravenous injection. Ten days after transplant, mice were treated with vehicle or 50 mg kg⁻¹ 6-MP per day for 5 days, then allowed to recover for ten days and given a second round of treatment for 1–3 days. Following this second cycle of treatment, mice were euthanized and lymphoblasts were recovered from spleen samples for quantitative evaluation of *Nt5c2^{+/R367Q}* mutant cells.

We generated primary human leukaemia xenografts by intravenous injection of cryopreserved leukaemia lymphoblasts from diagnostic and relapsed acute lymphoblastic leukaemia patient samples into immunodeficient NRG mice.

We infected primary leukaemia xenograft cells with lentiviral particles expressing the red cherry fluorescent protein and luciferase (FUW-mCherry-Luc-puro) and transplanted matched ALL-4 diagnosis and ALL-4 relapse tumour cells into NRG immunodeficient recipients by intravenous injection and monitored tumour development by *in vivo* luminescence bioimaging with the *In vivo* Imaging System (IVIS, Xenogen) and by analysis of human CD45⁺ cells in peripheral blood by flow cytometry with an APC conjugated antibody (eBioscience 17-0459-42). Upon tumour establishment, mice were treated intraperitoneally with PBS vehicle or mizoribine (100 mg kg⁻¹ administered twice a day) for 3 consecutive days. Four animals in the relapse-xenograft mizoribine treatment group did not tolerate the full course of therapy presumably because of tumour lysis syndrome and were not included in the analysis. We euthanized mice on day 4 and analysed spleen weight and CD45⁺ tumour infiltration in the bone marrow by flow cytometry.

***In vitro* cell viability and chemotherapy response assays.** We measured cell growth and chemotherapy responses of primary mouse tumours, patient-derived xenografts, and human ALL cell lines *in vitro* by measurement of the metabolic reduction of the tetrazolium salt MTT using the Cell Proliferation Kit I (Roche) following the manufacturer's instructions. We analysed chemotherapy responses following 72-h incubation with increasing concentrations of 6-mercaptopurine or mizoribine.

For the mixed culture experiment of isogenic wild-type *Nt5c2^{+/co-R367Q}* and *Nt5c2^{+/R367Q}* mouse tumour cells, we treated uninfected tumour cells (expressing GFP) with vehicle and treated the same tumour cells previously infected with a mCherry-expressing vector (FUW-mCherry-Luc-puro) with 4-hydroxytamoxifen and quantified proportions of the two cell populations by FACS analysis using a Fortessa flow cytometer (BD Biosciences) and analysed data with FlowJo software (FlowJo LLC). All experiments were performed in triplicate.

Cell synchronization and cell cycle analysis. We synchronized isogenic wild-type *Nt5c2^{+/co-R367Q}* and *Nt5c2^{+/R367Q}* mouse tumour cells using a double thymidine block procedure. In brief, we incubated cells with 2 mM thymidine (Sigma Aldrich) for 16 h, allowed cells to recover for 14 h in regular media, and incubated a second time with 2 mM thymidine for 16 h. We harvested cells at 0, 3, 6 and 9 h time points and stained them with propidium iodide (Sigma Aldrich) for cell cycle progression analysis. FACS analysis was performed using a FACSCanto flow cytometer (BD Biosciences) and we analysed data with FlowJo software (FlowJo LLC).

Quantitative allele-specific qPCR assay. We quantitatively assessed the presence of the allele encoding NT5C2(R367Q) in matching DNA specimens obtained at diagnosis and during remission using a custom Mutation Detection Competitive Allele-Specific TaqMan PCR (castPCR) Assay (Life Technologies) using 30 ng

of DNA in a reaction volume of 20 μ l in a 7500 real-time PCR system (Applied Biosystems) following the manufacturer's instructions and recommended cycling conditions. We determined a detection ΔC_t cut-off value for the assay by running the wild-type and mutant *NT5C2* assays on genomic DNA samples from three wild-type cell lines and calibrated both assays by spiking in increasing concentrations of wild-type *NT5C2* or of the plasmid containing the *NT5C2* allele encoding the R367Q mutation. We determined the assay sensitivity for the allele encoding *NT5C2*(R367Q) by analysing *NT5C2* wild-type genomic DNA samples spiked with decreasing concentrations of the plasmid containing the *NT5C2* allele encoding the R367Q mutation. We analysed experimental data using the Mutation Detector Software (Life Technologies) to calculate the ΔC_t value between the wild-type *NT5C2* and the *NT5C2*(R367Q)-encoding allele assay reads for each sample, and comparing these to the predetermined ΔC_t cut-off value.

To quantitatively assess the presence of *Nt5c2^{+/R367Q}* mutant cells in mixed tumour populations of wild-type *Nt5c2^{+/co-R367Q}* and *Nt5c2^{+/R367Q}* mutant lymphoblasts, we performed a quantitative analyses of mutant transcripts normalizing tumour content by quantitative PCR with reverse transcription (RT-PCR) analysis of GFP. In this experiment we isolated RNA from lymphoblasts with the RNeasy kit (Qiagen) and prepared complementary DNA (cDNA) by reverse transcription using the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen). *Nt5c2* exon 14 was amplified using TaqMan Gene Expression Master Mix (TaqMan) and the allele encoding *Nt5c2*(R367Q) was detected using a mutant-specific TaqMan probe (5' FAM-AGGGTGGCAGACTTT-MGBNFQ 3', ThermoFisher). *Actb* (β -actin) and GFP were amplified using FastStart Universal SYBR Green Master (ROX) (Roche) following standard protocols. Quantitative PCR reactions were run in a 7500 Real Time PCR System (Applied Biosystems). C_t values of the allele encoding *Nt5c2*(R367Q) and GFP were normalized to *Actb* C_t values and a ratio was taken of *Nt5c2*(R367Q) expression over GFP expression to represent the percentage of *Nt5c2^{+/R367Q}* mutant cells present in mixed tumour populations.

Digital droplet PCR. Targeted ultra-deep mutation screening was performed using the digital droplet PCR technique (RainDance Technologies) as described previously³³. In brief, TaqMan assay primers and probes were custom designed for the allele encoding NT5C2(P414A) using PrimerExpress 3.0 (Thermo Fisher Scientific). Primers and probes for the allele encoding NT5C2(R39Q) were designed through the Custom TaqMan Assay Design Tool (CADT) (Life Technologies) with the support of RainDance Technologies. Probes matching the wild-type allele were labelled with VIC fluorescent reporter dye and probes matching the mutant allele were labelled with FAM dye. Amplicon sizes ranged from 75 bp to 120 bp. Genomic DNA was sheared to 3 kb using the M220 instrument (Covaris) and a total of 500–1,000 ng of fragmented DNA was used in each 50 μ l ddPCR reaction. The digital droplet PCR reaction further contained 1 \times TaqMan Genotyping Master Mix (Applied Biosystems), 1 \times digital PCR droplet stabilizer (RainDance Technologies), and 1 \times TaqMan primers and probes mix (Integrated DNA Technologies). In line with the manufacturer's instructions, an average of 7×10^6 droplets were generated by the RainDrop Source instrument and emulsion PCR was performed using the C1000 Thermal Cycler (BioRad). Droplet fluorescence of the amplified product was detected by the RainDrop Sense instrument and data analysis was carried out using the RainDrop Analyst II Software (RainDance Technologies).

To determine the detection limit of the assays, we constructed dilution curves of patient tumour cells and cells from the REH cell line. The REH cell line was confirmed to be wild type after Sanger sequencing for the locations targeted in the digital droplet PCR. We collected pure populations of tumour cells by flow cytometric sorting of the relapse samples of patients SJBALL192 (containing cells heterozygous for the allele encoding NT5C2(R39Q)). We made serial dilutions of tumour cells with wild-type cells (REH cell line) to generate final mutant allele frequency (MAF) levels of 50%, 5%, 0.5%, 0.05%, 0.005% and 0.0005% and isolated DNA using phenol–chloroform. With an input of 500 ng DNA in the digital droplet PCR assay the MAFs correspond to 70,000, 7,000, 700, 70, 7, and 0.7 copies, respectively. A frequency of >0.005% (>7 copies) could be consistently detected.

Duplex sequencing of diagnostic patient samples. Duplex sequencing was carried out by TwinStrand Biosciences under fully blinded conditions using methods previously described^{34,35}. In brief, 400 ng of extracted genomic DNA was ultrasonically sheared, A-tailed and ligated to degenerate tag-containing Duplex adapters. The library was amplified and subjected to two successive rounds of hybrid capture with 120 bp biotinylated oligonucleotide probes tiled across exons 9, 11, 13, 15, 16 and 17 of the human *NT5C2* gene and flanking sequences. Indexed libraries were pooled and sequenced on an Illumina NextSeq 500. Duplex consensus sequences were generated after alignment to hg38 using the requirement that error-corrected bases be supported by at least three independent reads from each original strand. The variant calls for each sample were filtered against known single nucleotide polymorphisms in the Phase 3 build of the 1000 genomes database and tabulated

versus all reference base calls at the eight codons of interest. Variant allele frequency was calculated as the number of variants per total number of error-corrected bases at each nucleotide position. The average error-corrected molecular depth at codons of interest was approximately $4,100 \times (1,840\text{--}8,530 \times)$, yielding an average power for detecting variants at a level of 1/1,000 of ~98%.

Metabolomic analyses. To analyse metabolic differences between *Nt5c2^{+/co-R367Q}* and *Nt5c2^{+/R367Q}* primary mouse tumours, we treated tumour cells in triplicate with 1 μ M 4-hydroxytamoxifen for 48 h *in vitro* to induce expression of the allele encoding Nt5c2(R367Q) or with vehicle for wild-type controls, after which we diluted out the 4-hydroxytamoxifen or vehicle with media. After 72 h, we harvested cells into packed 50–100- μ l size pellets and collected conditioned media for analyses ($n = 3$, cell pellets and media). We flash-froze cell pellet and media samples, which were then extracted using standard solvent extraction methods and analysed on the gas chromatography–mass spectrometry and liquid chromatography–tandem mass spectrometry platforms by Metabolon. Analysed metabolites consisted of a total of 459 named biochemicals in cells and 252 named biochemicals in media. We first normalized results to protein concentration, log transformed and imputed any missing values with the minimum observed value for each compound. We then used Welch's two-sample *t*-test to identify biochemicals that differed significantly between experimental groups. To account for the multiple comparisons that occur in metabolomics studies we also calculated an estimate of the false discovery rate (*q*-value), which indicates the fraction of biochemicals that would meet a given *P*-value cut-off by random chance. Similar processing and analyses were performed on CUTLL1 and REH ALL cell lines expressing wild-type NT5C2 or NT5C2(R367Q). Analysed metabolites in these cell lines consisted of a total of 596 named biochemicals in cells and 347 in media.

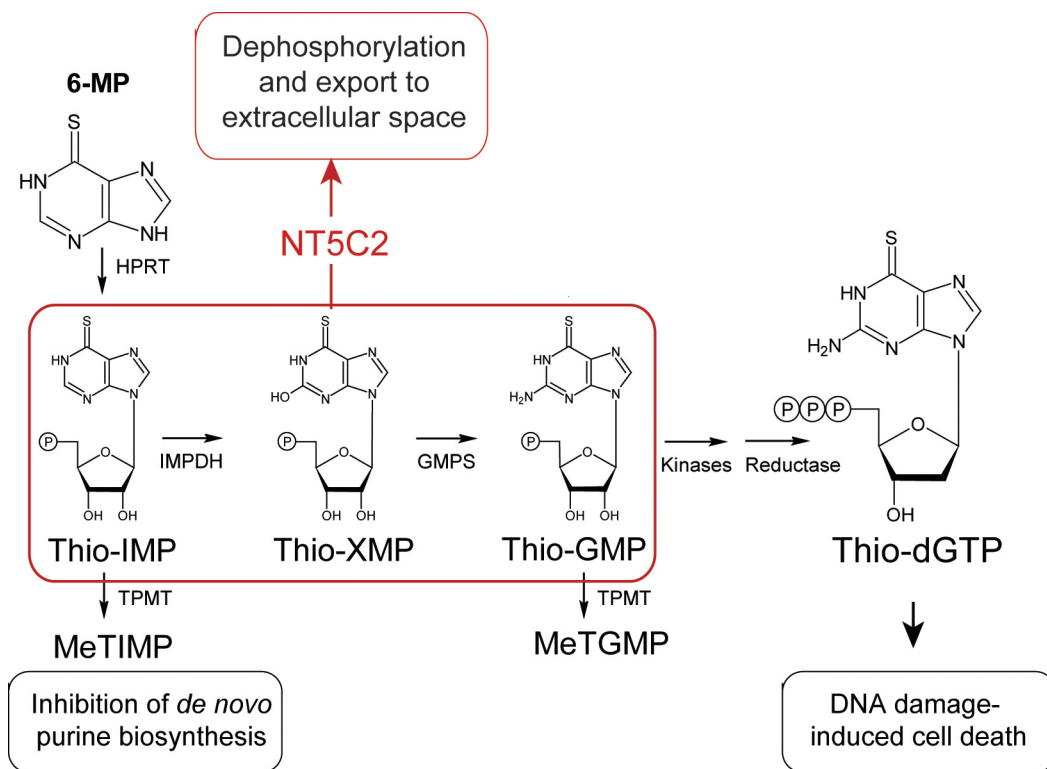
ISN of relapsed ALL. We illustrated the sequential order of somatic mutations in relapsed ALL using the ISN²⁵ that pools evolutionary paths across all patients. We selected recurrently mutated genes that were previously defined as drivers of paediatric ALL^{36–38} and relapse-genes^{10,11}. Only non-synonymous single nucleotide variants were used in analysis. For each patient, we generated a sequential network that defined early events as mutations observed in both the primary tumour and the relapsed tumour, whereas late events were mutations only observed in the relapsed tumour. Each node represented a gene, and each arrow pointed from a gene with an early event to a gene with a late event. The ISN then pooled sequential networks across all patients. To test whether a gene within the ISN was significantly

early or late, we used the binomial test based on the in-degree and out-degree of each node. Somatic mutation data used to generate ISN were aggregated from previously published studies (refs 9–11).

Statistical analyses. We performed statistical analysis by Student's *t*-test. We considered results with $P < 0.05$ as statistically significant. Survival in mouse experiments was represented with Kaplan–Meier curves and significance was estimated with the log-rank test (GraphPad Prism). We analysed serial limited dilution leukaemia-initiating cell data using the ELDA software³⁹. No outlier data points were excluded in the analyses.

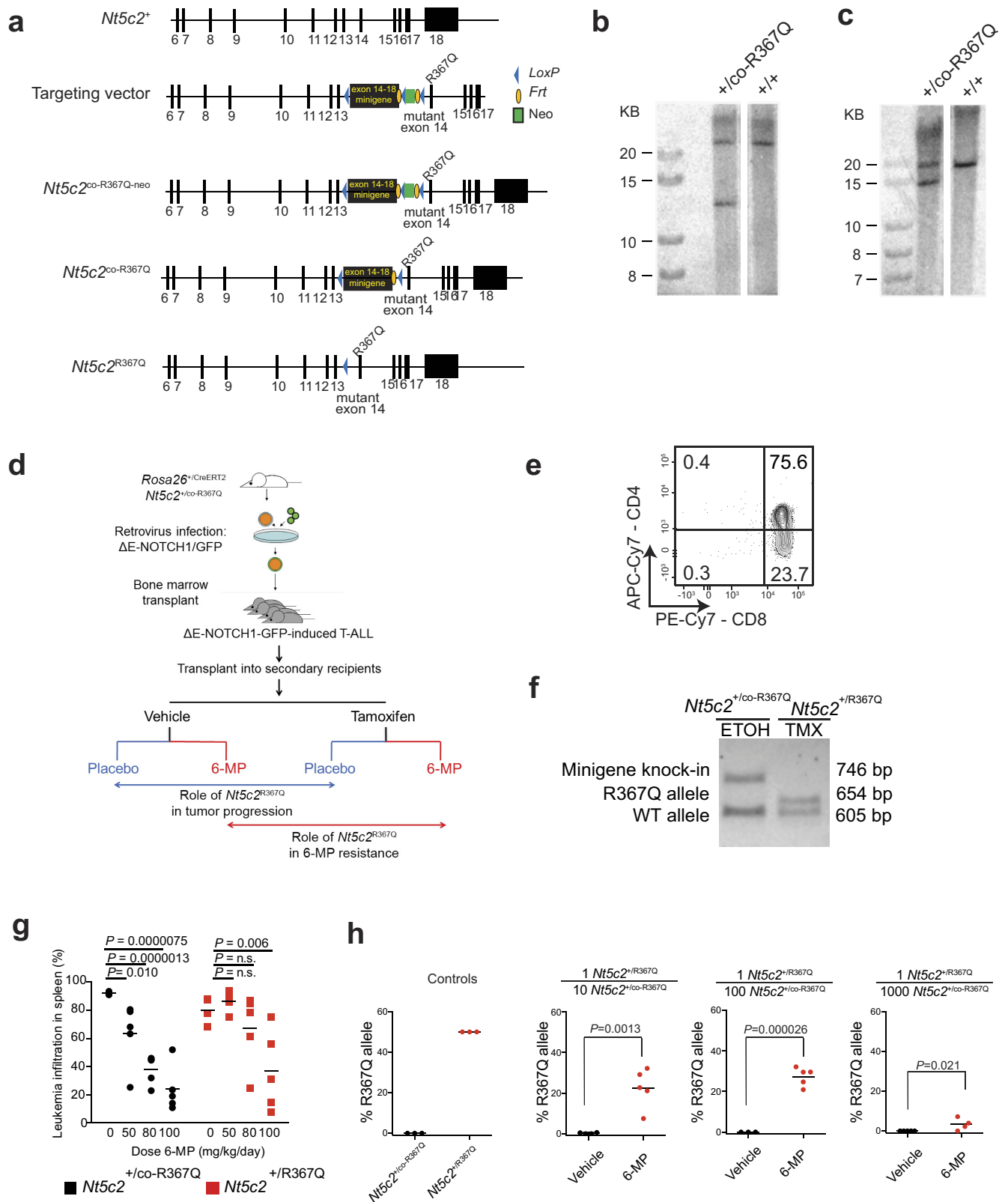
Data availability. All data generated or analysed during this study are included in this published article and its Supplementary Information.

29. Palomero, T. *et al.* CUTLL1, a novel human T-cell lymphoma cell line with t(7;9) rearrangement, aberrant NOTCH1 activation and high sensitivity to gamma-secretase inhibitors. *Leukemia* **20**, 1279–1287 (2006).
30. Kimbrel, E. A. *et al.* Systematic *in vivo* structure-function analysis of p300 in hematopoiesis. *Blood* **114**, 4804–4812 (2009).
31. Guo, K. *et al.* Disruption of peripheral leptin signaling in mice results in hyperleptinemia without associated metabolic abnormalities. *Endocrinology* **148**, 3987–3997 (2007).
32. Herranz, D. *et al.* Metabolic reprogramming induces resistance to anti-NOTCH1 therapies in T cell acute lymphoblastic leukemia. *Nat. Med.* **21**, 1182–1189 (2015).
33. Iacobucci, I. *et al.* Truncating erythropoietin receptor rearrangements in acute lymphoblastic leukemia. *Cancer Cell* **29**, 186–200 (2016).
34. Schmitt, M. W. *et al.* Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl Acad. Sci. USA* **109**, 14508–14513 (2012).
35. Kennedy, S. R. *et al.* Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat. Protoc.* **9**, 2586–2606 (2014).
36. Futreal, P. A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
37. Van Vlierberghe, P. & Ferrando, A. The molecular basis of T cell acute lymphoblastic leukemia. *J. Clin. Invest.* **122**, 3398–3406 (2012).
38. Zhang, J. *et al.* Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. *Blood* **118**, 3080–3087 (2011).
39. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).



Extended Data Figure 1 | Schematic representation of 6-MP activation and mechanism of action. The hypoxanthine-guanine phosphoribosyl transferase enzyme (HPRT) processes 6-MP to thio-IMP, which is then converted to thio-XMP and thio-GMP. Subsequent metabolism of thio-GMP by kinases and reductases yields thio-dGTP which is incorporated into replicating DNA strands and triggers the DNA mismatch-repair

machinery, leading to cell cycle arrest and apoptosis. The anti-leukaemic effects of 6-MP are in part also attributed to a second metabolic pathway in which thiopurine S-methyl transferase (TPMT) methylates thio-IMP to form methylthio-IMP (MeTIMP), which is a potent inhibitor of amidophosphoribosyltransferase (ATase), an enzyme catalysing the committed step of *de novo* purine biosynthesis.

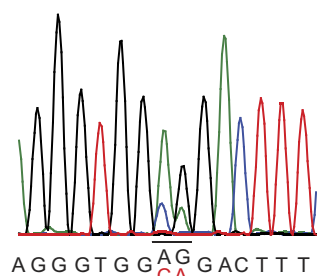


Extended Data Figure 2 | See next page for caption.

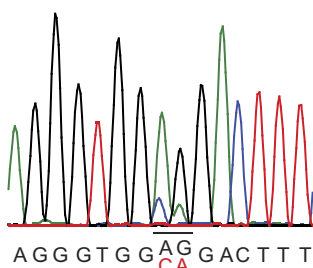
Extended Data Figure 2 | Conditional knock-in targeting of *Nt5c2*, generation and analysis of a *Nt5c2*(R367Q) conditional inducible T-ALL model. **a**, Schematic representation of the targeting strategy for the generation of *Nt5c2*^{+/co-R367Q} conditional knock-in mice. **b**, Southern blot analysis of DNA samples from *Nt5c2*^{+/+} and targeted *Nt5c2*^{+/co-R367Q} embryonic stem cells after digestions with BamHI restriction enzyme and hybridization of a DNA probe external to the long arm. **c**, Southern blot analysis of DNA samples from *Nt5c2*^{+/+} and targeted *Nt5c2*^{+/co-R367Q} embryonic stem cells after digestion with ApaI restriction enzyme and hybridization of a DNA probe to the short arm. **d**, Schematic depiction of the strategy for developing conditional inducible *Nt5c2*^{+/co-R367Q} primary mouse T-ALL tumours and for assessing the role of *Nt5c2*^{+/R367Q} on leukaemia progression and response to chemotherapy. **e**, Representative FACS plot of a *Rosa26*^{+/creERT2}*Nt5c2*^{+/co-R367Q} ΔE-NOTCH1-induced primary T-ALL tumour with a CD4⁺CD8⁺ immunophenotype.

f, Representative genotyping PCR results from genomic DNA of a *Rosa26*^{+/CreERT2}*Nt5c2*^{+/co-R367Q} ΔE-NOTCH1-induced primary T-ALL tumour treated with 4-hydroxytamoxifen (TMX) or vehicle only (ethanol, ETOH) *in vitro* showing Cre-mediated deletion of the exon 14–18 *Nt5c2* wild-type mini-gene. **g**, Tumour burden assessed in the spleen (percentage of GFP⁺ cells) in mice allografted with *NOTCH1*-induced *Nt5c2*^{+/co-R367Q} and isogenic *Nt5c2*^{+/R367Q} primary leukaemia cells treated with a range of 6-MP doses (*n* = 5 per group). **h**, Analysis of selection for the mutant allele encoding *Nt5c2*(R367Q) by qPCR in mice allografted with *Nt5c2*^{+/co-R367Q} and *Nt5c2*^{+/R367Q} primary mouse T-ALL cells at a 1:10, 1:100 and 1:1,000 *Nt5c2*^{+/R367Q}:*Nt5c2*^{+/co-R367Q} dilution and treated with vehicle or 6-MP (*n* = 5 mice per group and *n* = 3 technical replicates for the controls). The horizontal bar represents mean values. *P* values were calculated using two-tailed Student's *t*-test (**g**) or a one-tailed Student's *t*-test (**h**). Data in **e** and **f** show representative results from more than two experiments.

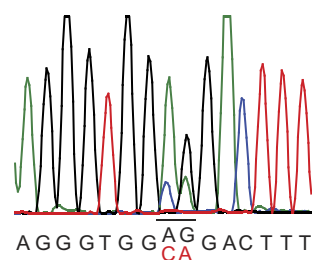
Mouse 1



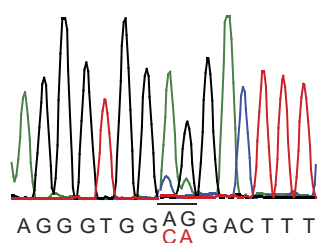
Mouse 2



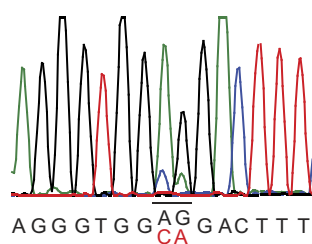
Mouse 3



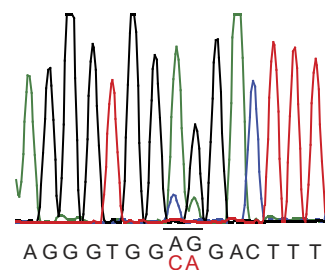
Mouse 4



Mouse 5

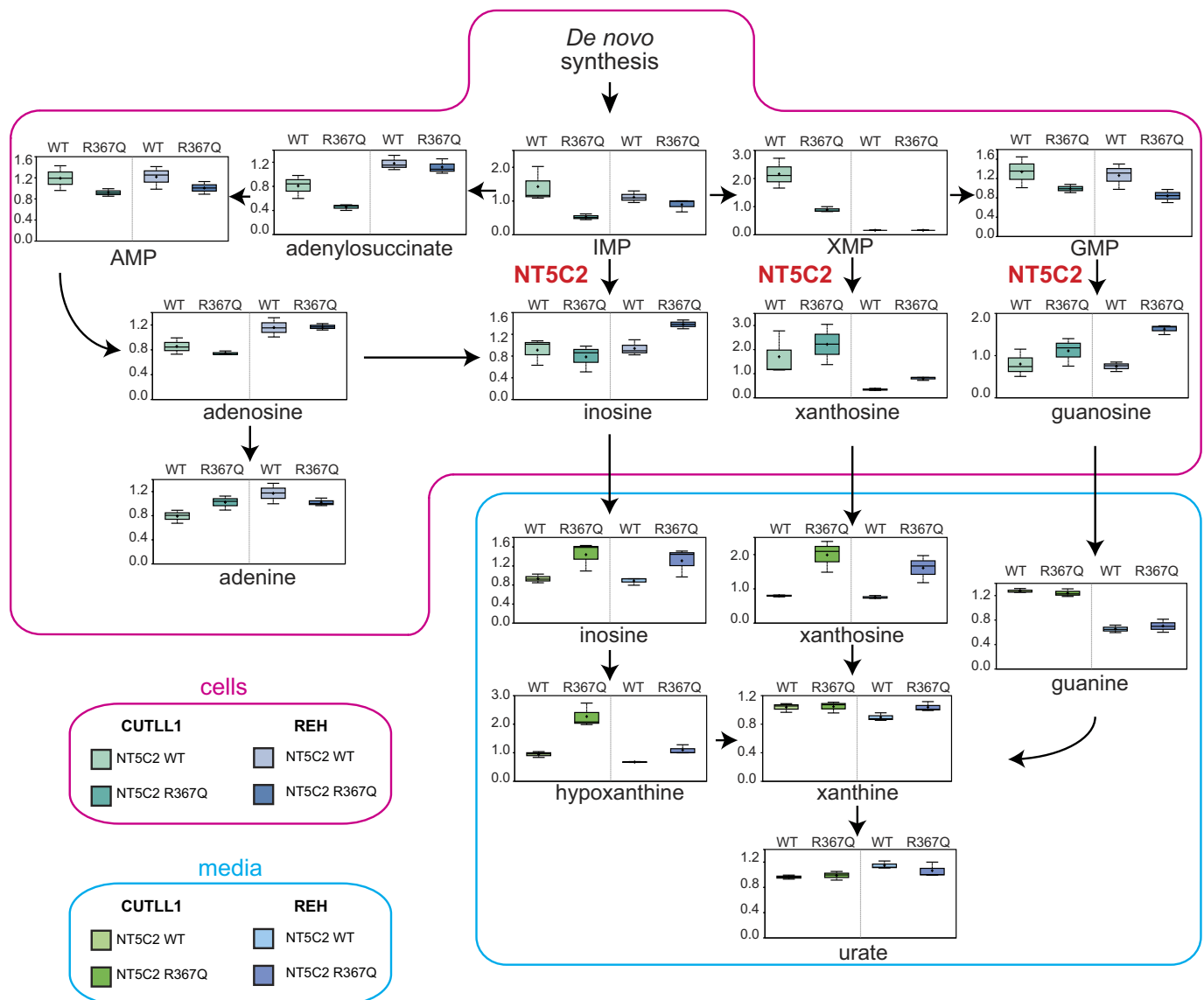


Mouse 6



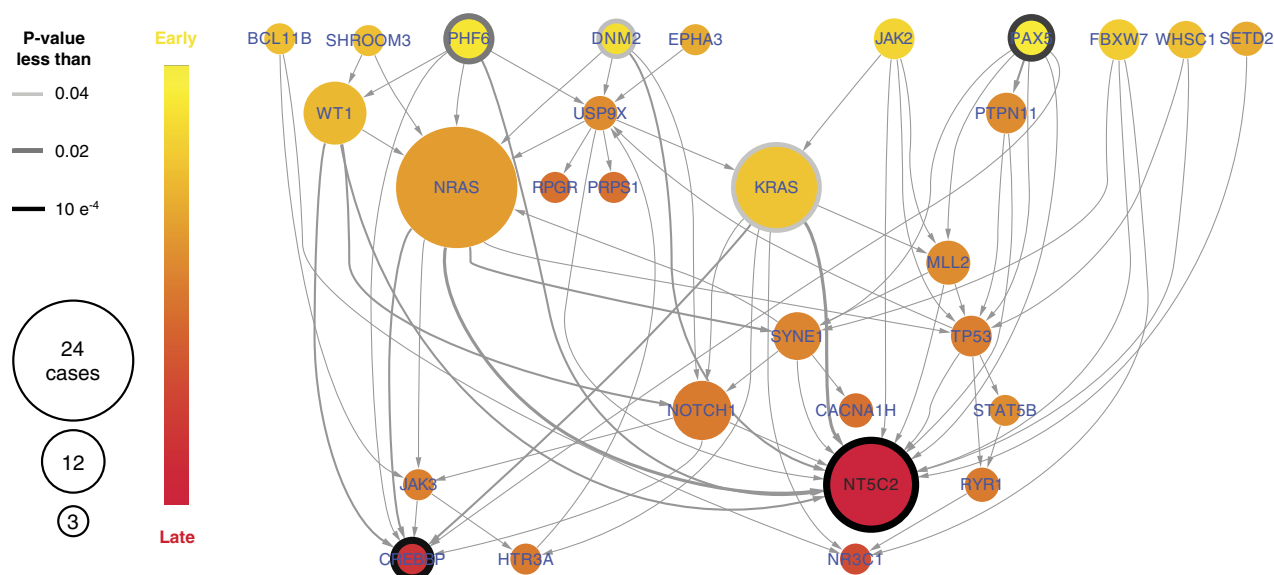
Extended Data Figure 3 | Decreased expression of the allele encoding Nt5c2(R367Q) allele upon leukaemia progression *in vivo*. Sanger sequencing chromatograms of cDNA from tumours in Fig. 2c show decreased expression of the Nt5c2(R367Q)-encoding allele over the

wild-type *Nt5c2* allele compared with recently 4-hydroxytamoxifen treated *Rosa26^{+/creERT2}Nt5c2^{+/co-R367Q}* cells (Fig. 1a). Mutant-allele deoxynucleotides are indicated in red.



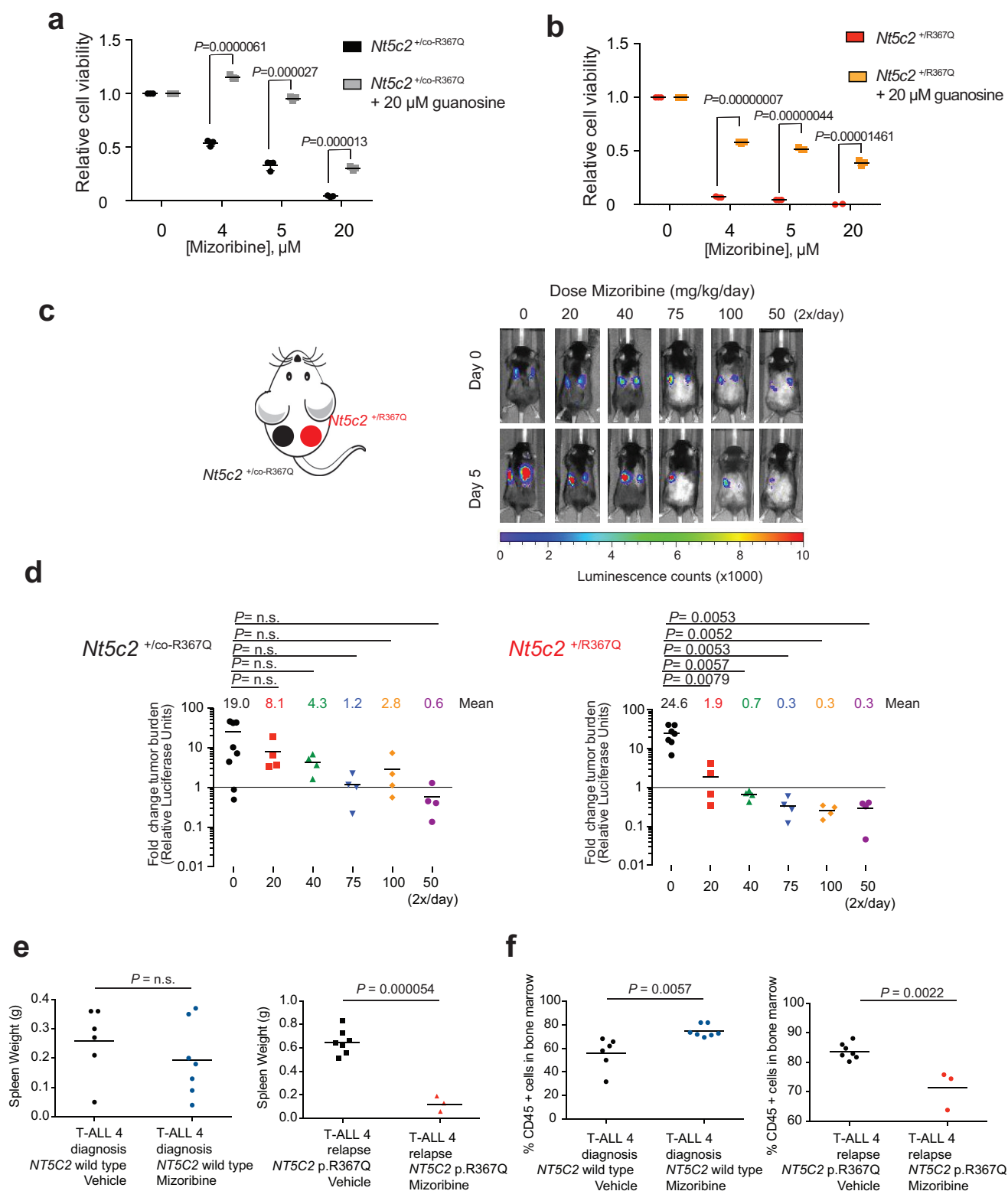
Extended Data Figure 4 | NT5C2(R367Q) expression leads to increased purine export in T-ALL and B-ALL cell lines. Diagram of the purine *de novo* biosynthesis and salvage pathways, showing gas chromatography–mass spectrometry and liquid chromatography–tandem mass spectrometry metabolic profiles (mass spectrometry scaled intensity, arbitrary units) of CUTLL1 and REH cell lines expressing

wild-type NT5C2 or NT5C2(R367Q) and their corresponding conditioned media ($n = 3$ biological replicates per sample). Box plots represent the upper quartile to lower quartile distribution. Plus signs indicate mean values, horizontal lines indicate median values and whiskers indicate the maximum and minimum values of the distributions.



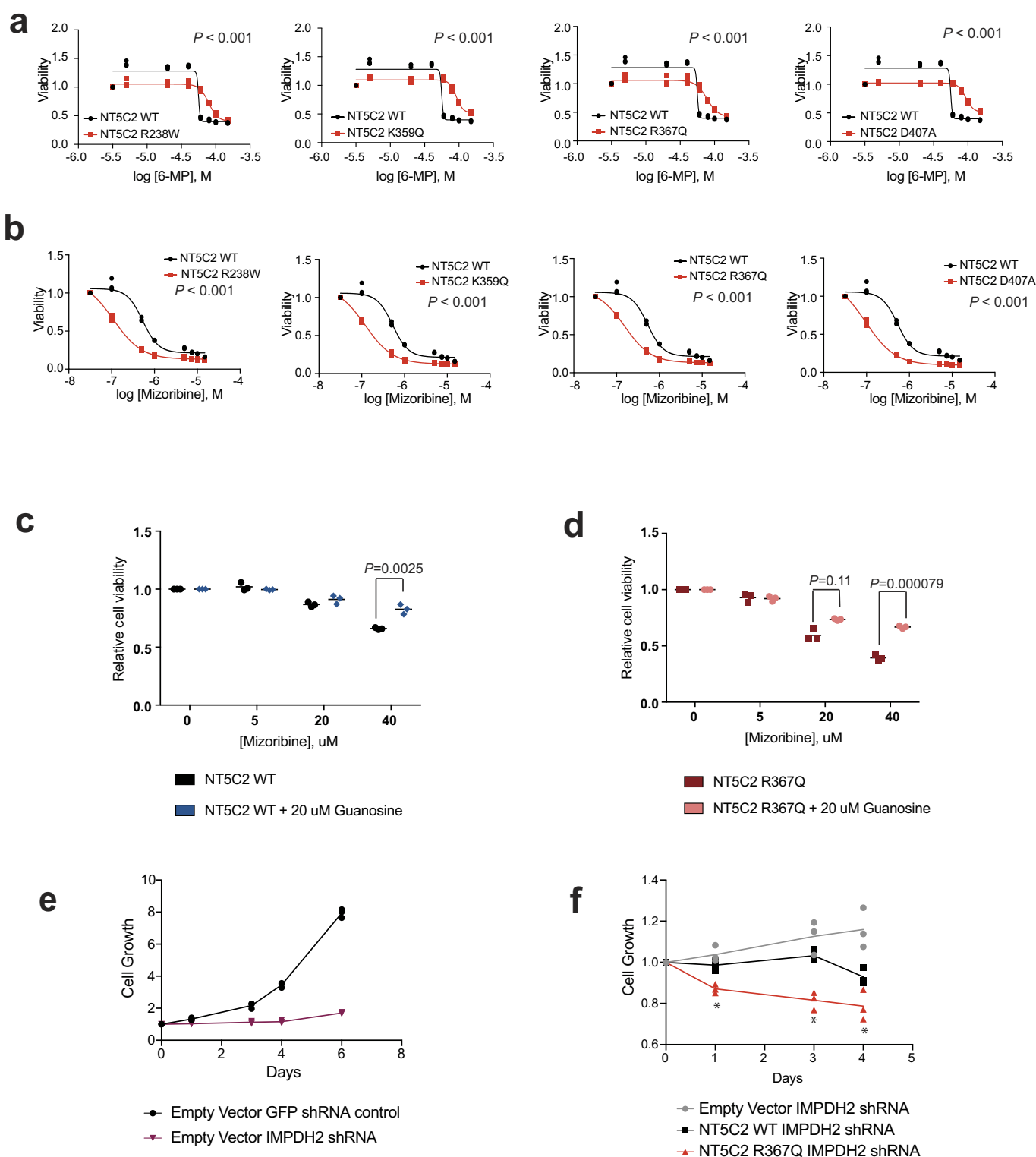
Extended Data Figure 5 | *NT5C2* mutations are late events in ALL. ISN illustrating the sequential order of somatic mutations in relapsed ALL by pooling evolutionary paths across patients. Each node represents a gene and each arrow points from a gene with an early event to a gene with a late

event. To test whether a gene within the ISN was significantly early or late, we used a one-sided binomial test based on the in-degree and out-degree of each node.



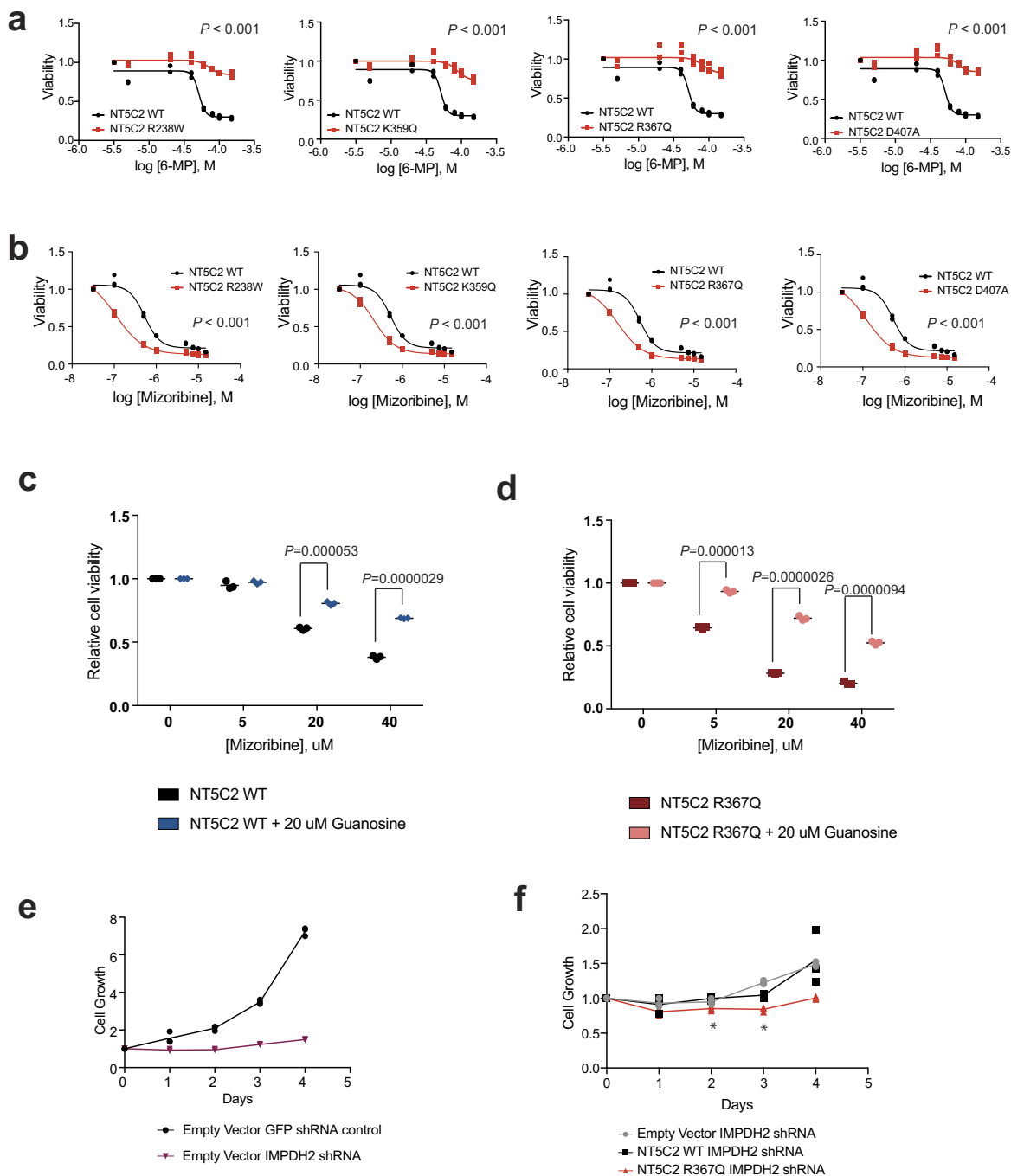
Extended Data Figure 6 | Guanosine rescue of mizoribine sensitivity *in vitro* and mizoribine activity against NT5C2(R367Q) mutant cells *in vivo*. **a, b,** Cell viability assays showing drug responses of wild-type $Nt5c2^{+/co-R367Q}$ primary mouse T-ALL cells (**a**) or mutant $Nt5c2^{+/R367Q}$ mouse T-ALL lymphoblasts (**b**) to increasing doses of mizoribine in the presence of $20 \mu\text{M}$ guanosine ($n=3$ biological replicates). **c,** Analysis of tumour burden assessed by bioimaging in mice transplanted with wild-type $Nt5c2^{+/co-R367Q}$ leukaemia cells (left flank) or mutant $Nt5c2^{+/R367Q}$ leukaemia cells (right flank) treated with a range of mizoribine doses ($n=8$ mice for the vehicle group and $n=4$ mice per treated group). **d,** Quantification of data from **c**. **e,** Analysis of tumour burden assessed

by spleen weight in mice allografted with wild-type NT5C2 ALL-4 diagnosis or NT5C2(R367Q) ALL-4 relapsed-patient derived leukaemia cells treated with 100 mg kg^{-1} mizoribine twice a day ($n=6$ for diagnosis vehicle group, $n=3$ for relapse treated group and $n=7$ for diagnosis treated and relapse vehicle groups). **f,** Analysis of tumour burden assessed by percentage of CD45⁺ cells in the bone marrow of mice allografted with NT5C2 wild-type ALL-4 diagnosis or NT5C2(R367Q) ALL-4 relapsed-patient derived leukaemia cells treated with 100 mg kg^{-1} mizoribine twice daily. ($n=3-7$ mice per group). Horizontal bars in **a, b, d-f** indicate mean values. P values were calculated using a two-tailed Student's t -test.



Extended Data Figure 7 | 6-MP and IMPDH inhibition response in CUTLL1 cells. **a**, Cell viability assays showing drug responses of the CUTLL1 cell line infected with wild-type or mutant *NT5C2*-expressing lentiviruses to increasing doses of 6-MP. **b**, Cell viability assays as in **a** documenting the response to mizoribine. **c**, **d**, Cell viability assay showing drug responses of wild-type (**c**) or *NT5C2*(R367Q) (**d**) CUTLL1 cells to increasing doses of mizoribine in the presence of 20 μ M guanosine.

e, Growth curve of CUTLL1 cells infected with a control short hairpin RNA (shRNA) targeting GFP or an shRNA targeting *IMPDH2*. **f**, Growth curve of wild-type or *NT5C2*(R367Q) CUTLL1 cells and infected with a shRNA targeting *IMPDH2*. **a–f**, Data are from three biological replicates. Horizontal bars in **c** and **d** indicate mean values. *P* values were calculated using a two-tailed Student's *t*-test. **P* \leq 0.05.



Extended Data Figure 8 | 6-MP and IMPDH inhibition response in REH B-ALL cells. **a**, Cell viability assay showing drug responses of the REH cell line infected with wild-type or mutant NT5C2-expressing lentiviruses to increasing doses of 6-MP. **b**, Cell viability assays as in **a** documenting the response to mizoribine. **c**, **d**, Cell viability assay showing drug responses of wild-type (**c**) or NT5C2(R367Q) (**d**) REH cells to increasing doses of

mizoribine in the presence of 20 μ M guanosine. **e**, Growth curve of REH cells infected with a control shRNA targeting GFP or shRNA targeting *IMPDH2*. **f**, Growth curve of wild-type or NT5C2(R367Q) REH cells and infected with an shRNA targeting *IMPDH2*. **a–f**, $n = 3$ biological replicates. Horizontal bars in **c** and **d** indicate mean values. P values were calculated using a two-tailed Student's t -test. * $P \leq 0.05$.

Extended Data Table 1 | Deep sequencing, allele-specific PCR and droplet PCR analyses of matched diagnostic and remission DNA from patients with *NT5C2* mutations at relapse

Duplex Sequencing			
Sample	Relapse Mutation	Average duplex depth	Allele load at diagnosis
T-ALL I1	K359Q	5009	Not detected
T-ALL 22	R238W	6519	Not detected
T-ALL 29	R391	3728	Not detected
T-ALL 30	Q523*	3512	Not detected
T-ALL C4	R367Q	5863	Not detected
T-ALL C5	R238W	3341	Not detected
T-ALL C7	R367Q	5134	Not detected
T-ALL C10	R238W	4299	Not detected
T-ALL C11	R367Q	3086	Not detected
T-ALL C14	D407E	3755	Not detected
T-ALL C17	R367Q	3497	Not detected
T-ALL C18	R367Q	3660	Not detected
T-ALL C20	R367Q	3693	Not detected
T-ALL N1	R367Q	3507	Not detected

NT5C2(R367Q) Allele Specific PCR			
Sample	Detection threshold	Allele load at diagnosis	Allele load at remission
T-ALL 4	1/1000	Not detected	-
T-ALL 17	1/1000	Not detected	Not detected
T-ALL 35	1/1000	Not detected	-
T-ALL C4	1/1000	Not detected	Not detected
T-ALL C7	1/1000	Not detected	Not detected
T-ALL C11	1/1000	Not detected	Not detected
T-ALL C17	1/1000	Not detected	Not detected
T-ALL C18	1/1000	Not detected	Not detected
T-ALL C20	1/1000	Not detected	Not detected

Serial Patient Sample Droplet PCR					
Sample	Sample Type	Days since diagnosis	Detection threshold (%)	Mutation	Mutant Allele Frequency (%)
SJBALL192	D	0	0.005	P414A	0.00000
SJBALL192	R1	170	0.005	P414A	37.73843
SJBALL192	CR	204	0.005	P414A	0.20425
SJBALL192	CR	230	0.005	P414A	0.15196
SJBALL192	R2	280	0.005	P414A	0.00064
SJBALL192	D	0	0.005	R39Q	0.00256
SJBALL192	R1	170	0.005	R39Q	0.00584
SJBALL192	CR	230	0.005	R39Q	0.02241
SJBALL192	R2	280	0.005	R39Q	48.57407
SJTALL001	D	0	0.005	R39Q	0.00238
SJTALL001	D	0	0.005	R39Q	0.00307
SJTALL001	CR	53	0.005	R39Q	0.00576
SJTALL001	CR	53	0.005	R39Q	0.00762
SJTALL001	CR	218	0.005	R39Q	0.00333
SJTALL001	CR	218	0.005	R39Q	0.00579
SJTALL001	CR	362	0.005	R39Q	0.01952
SJTALL001	R1	399	0.005	R39Q	3.70027
SJTALL001	R1	412	0.005	R39Q	3.57648
SJTALL001	CR	434	0.005	R39Q	0.00931
SJTALL001	R2	751	0.005	R39Q	37.58709
SJTALL001	R2	751	0.005	R39Q	39.23768

D = Diagnosis, CR=complete remission, R1=First Relapse, R2=Second Relapse

Extended Data Table 2 | Leukaemia-initiating cell activity of isogenic *Nt5c2*^{+/co-R367Q} wild-type and *Nt5c2*^{+/R367Q} primary mouse T-ALL tumours

Rosa26^{+/CreERT2} <i>Nt5c2</i> R367Q^{+/co-R367Q} – Vehicle treated		
Number of cells injected / mouse	Number of mice injected	Number of leukemia-developing mice
100000	6	6
10000	6	6
1000	6	5
100	6	2
10	6	0

Rosa26^{+/CreERT2} <i>Nt5c2</i> R367Q^{+/R367Q} – Tamoxifen treated		
Number of cells injected / mouse	Number of mice injected	Number of leukemia-developing mice
100000	6	6
10000	6	3
1000	6	3
100	5	0
10	5	0

A *Myc* enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies

Carsten Bahr^{1,2,3*}, Lisa von Paleske^{1,2,3*}, Veli V. Uslu^{4*}, Silvia Remeseiro⁴, Naoya Takayama^{5,6}, Stanley W. Ng⁷, Alex Murison^{5,8}, Katja Langenfeld⁴, Massimo Petretich⁴, Roberta Scognamiglio^{1,2,3}, Petra Zeisberger^{1,2}, Amelie S. Benk^{1,2,3}, Ido Amit⁹, Peter W. Zandstra^{7†}, Mathieu Lupien^{5,8}, John E. Dick^{5,6}, Andreas Trumpp^{1,2,3,10,11§} & François Spitz^{4,12,13§}

The transcription factor *Myc* is essential for the regulation of haematopoietic stem cells and progenitors and has a critical function in haematopoietic malignancies¹. Here we show that an evolutionarily conserved region located 1.7 megabases downstream of the *Myc* gene that has previously been labelled as a ‘super-enhancer’² is essential for the regulation of *Myc* expression levels in both normal haematopoietic and leukaemic stem cell hierarchies in mice and humans. Deletion of this region in mice leads to a complete loss of *Myc* expression in haematopoietic stem cells and progenitors. This caused an accumulation of differentiation-arrested multipotent progenitors and loss of myeloid and B cells, mimicking the phenotype caused by *Mx1-Cre*-mediated conditional deletion of the *Myc* gene in haematopoietic stem cells³. This super-enhancer comprises multiple enhancer modules with selective activity that recruits a compendium of transcription factors, including GFI1b, RUNX1 and MYB. Analysis of mice carrying deletions of individual enhancer modules suggests that specific *Myc* expression levels throughout most of the haematopoietic hierarchy are controlled by the combinatorial and additive activity of individual enhancer modules, which collectively function as a ‘blood enhancer cluster’ (BENC). We show that BENC is also essential for the maintenance of MLL–AF9-driven leukaemia in mice. Furthermore, a BENC module, which controls *Myc* expression in mouse haematopoietic stem cells and progenitors, shows increased chromatin accessibility in human acute myeloid leukaemia stem cells compared to blasts. This difference correlates with *MYC* expression and patient outcome. We propose that clusters of enhancers, such as BENC, form highly combinatorial systems that allow precise control of gene expression across normal cellular hierarchies and which also can be hijacked in malignancies.

The *Myc* gene resides in a three-megabase (Mb)-long gene-poor region that corresponds to overlapping regulatory and topologically associating domains^{4,5}. Within this interval, multiple positions have chromatin features that are associated with cell-type-specific enhancer activities^{6–8}, and recent studies have shown that some of these distant elements have a role in the control of *Myc* expression^{5,9–12}. In haematopoietic tissues, this regulatory landscape appears to be complex. Enhancer-associated chromatin marks found in the bone marrow and fetal liver, but not in non-haematopoietic tissues, comprise two isolated centromeric peaks (450–550 kilobases (kb) upstream of *Myc*), a proximal cluster of peaks (from 50 kb to 450 kb downstream of *Myc*) that overlaps with the *Pvt1* gene, and another cluster of peaks located at the

distal end of the topologically associating domain, 1.7-Mb telomeric to *Myc* (Fig. 1a and Extended Data Fig. 1). We refer to this latter region as the blood enhancer cluster (BENC) (Fig. 1a and Extended Data Figs 1, 2a). Consistent with possible enhancer activity, the *Pvt1*-associated cluster and BENC have been shown to be in physical proximity to *Myc* in human and mouse blood cells^{2,13,14}. Both regions have been suggested to regulate *Myc* expression in mouse leukaemia cell lines^{2,15}, even though recent data have suggested a minimal role of BENC in human K562 leukaemia cells¹⁶. However, the contribution of these different regions to *Myc* expression *in vivo* remains unknown.

To study the functional activity of these different regions *in vivo*, we first took advantage of enhancer-sensors inserted into the locus⁵. A LacZ enhancer-sensor inserted close to BENC (sensor 17a) showed strong expression in haematopoietic stem cells (HSCs), multipotent progenitors, haematopoietic stem and progenitor cells (Lin[–]Sca-1⁺Kit⁺ (LSK)) and lineage-committed progenitors (Lin[–]Sca-1[–]Kit⁺ (LS[–]K)), as well as common myeloid progenitors and granulocyte-monocyte progenitors but not in differentiated mature cells (Lin⁺) (Fig. 1b and Extended Data Fig. 2b–k). Haematopoietic expression in adult mice was not observed when the inserted sensor was located elsewhere, including at location 3a, next to the *Pvt1* enhancer-associated peaks (Fig. 1a, b and Extended Data Fig. 2). Notably, three genomic deletions encompassing BENC that retain the LacZ sensor at 17a showed a complete loss of LacZ expression (Fig. 1a, b and Extended Data Fig. 2). Therefore, BENC appears to be the only enhancer region that is active in HSCs of adult mice *in vivo*.

We assessed the role of BENC *in vivo* by investigating the phenotypic consequences of its deletion. Mice homozygous for a deletion in the region between sensors 15a and 17a, which includes BENC (hereafter referred to as *Myc*^{Δ15–17/Δ15–17}), were viable and had a normal weight, but displayed a significant reduction in bone marrow cellularity (Extended Data Fig. 3a, b). Flow cytometric analysis revealed a significant increase in LSK cells, due to the robust accumulation of multipotent progenitor (MPP) populations 2, 3 and 4¹⁷, whereas HSC and MPP1 numbers remained unchanged (Fig. 1c, d and Extended Data Fig. 3c). In more mature compartments, we found an increased number of megakaryocytic cells, whereas granulocytes and B cells were severely depleted. By contrast, the number of T cells and erythrocytes was unaffected (Fig. 1e and Extended Data Fig. 3d). To functionally define the capacity of *Myc*^{Δ15–17/Δ15–17} HSCs, we performed competitive repopulation assays using an 80% donor:20% competitor ratio. In this context, *Myc*^{Δ15–17/Δ15–17} cells

¹Division of Stem Cells and Cancer, German Cancer Research Center (DKFZ) and DKFZ-ZMBH Alliance, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. ²Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM gGmbH), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. ³Faculty of Biosciences, University of Heidelberg, 69120 Heidelberg, Germany. ⁴Developmental Biology Unit, European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany. ⁵Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario M5G 2M9, Canada. ⁶Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5G 1A1, Canada. ⁷Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario M5G 1A1, Canada. ⁸Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 2M9, Canada. ⁹Department of Immunology, Weizmann Institute of Science, Rehovot 76100, Israel. ¹⁰German Cancer Consortium (DKTK), 69120 Heidelberg, Germany. ¹¹Nationales Zentrum für Tumorerkrankungen (NCT), 69120 Heidelberg, Germany. ¹²CNRS, UMR3738, 25 Rue du Dr Roux, 75015 Paris, France. ¹³(Epi)genomics of Animal Development Unit, Developmental and Stem Cell Biology Department, Institut Pasteur, 75015 Paris, France. [†]Present address: Michael Smith Laboratories, School of Biomedical Engineering, The University of British Columbia, #301 - 2185 East Mall, Vancouver, British Columbia V6T 1Z4, Canada.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

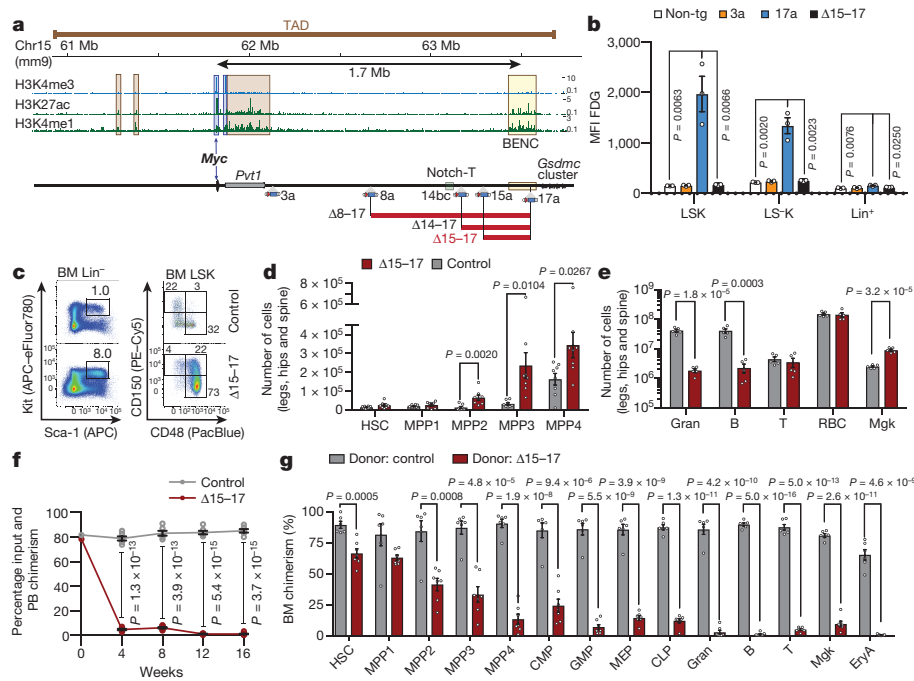


Figure 1 | The BENC enhancer region (15–17), located 1.7 Mb downstream of the *Myc* locus, is essential for HSC function. **a**, Top, schematic representation of the *Myc* locus within its topologically associating domain (TAD, brown bar)⁴, including the position of genes (arrows), predicted promoters (H3K4me3, blue boxes) and enhancers (H3K4me1, H3K27ac, other boxes). Bottom, representation of the different mouse transposon insertions (top) and Cre-mediated deletions (bottom, red bars) used to functionally characterize enhancer regions⁵. **b**, Quantitative analysis of LacZ activity as determined by fluorescein di-β-D-galactopyranoside (FDG) staining and flow cytometry of stem and progenitor cells (LSK), committed progenitors (LS⁺K) and differentiated cells (Lin⁺) isolated from wild-type (Non-tg), heterozygous 3a, 17a and Δ15–17 mice. MFI, geometric mean fluorescence intensity. **c–e**, Comparison of control and homozygous *Myc*^{Δ15–17/Δ15–17} mice.

showed little contribution to peripheral blood of recipient mice (Fig. 1f and Extended Data Fig. 3e). At 16 weeks after transplantation, *Myc*^{Δ15–17/Δ15–17} HSCs were present in the bone marrow of chimaeras, but in significantly reduced numbers (Fig. 1g and Extended Data Fig. 3f). However, compared to wild-type competitors, these mutant HSCs produced lower numbers of MPP2–4 and an even more reduced number of more differentiated progenitors, leading to the complete absence of mature differentiated cells including T cells and erythroid cells that are derived from these progenitors. These data demonstrate that, although *Myc*^{Δ15–17/Δ15–17} HSCs can engraft and show long-term persistence in recipient bone marrow, these cells have reduced self-renewal capacity and are defective in generating differentiated progenitors and mature cells, including T cells and erythroid cells, in a competitive setting. Importantly, *Myc*^{Δ15–17/Δ15–17} LS⁺K and LSK cells are able to efficiently colonize the thymus and generate mature T cells when transplanted in T-cell-deficient NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ (NSG) mice (Extended Data Fig. 3g, h). Taken together, these data show that BENC encodes a critical regulatory activity necessary for HSC and progenitor function. Without this regulatory input, these cells are either defective in their self-renewal, proliferation and differentiation potential or show a competitive disadvantage, as is the case for *Myc*^{Δ15–17/Δ15–17} lymphoid progenitors, the malfunction of which leads to loss of T cells.

The pattern of BENC activity and the phenotypic similarities between BENC mutants and *Mx1-cre* (hereafter *Mx-Cre*)-mediated conditional deletion of *Myc* (referred to as *Myc*^{ΔMx}) in adult bone marrow³ strongly suggest that BENC may be regulating *Myc* during haematopoiesis. In line with this hypothesis, chromatin conformation capture performed

c, Representative cell surface expression profiles of indicated markers of the stem and progenitor cell compartments by flow cytometry. **d**, Number of HSCs and MPP1–4. **e**, Number of mature cell types. Gran, granulocytes; Mfg, megakaryocytes; RBC, red blood cells. **f, g**, Transplantation of homozygous *Myc*^{Δ15–17/Δ15–17} bone marrow (BM) cells in a competitive setting. The level of chimaerism is shown in the peripheral blood (PB) (**f**) and indicated cell types in the bone marrow 16 weeks after transplantation (**g**). CMP, common myeloid progenitors; CLP, common lymphoid progenitor; EryA, erythrocyte lineage A; GMP, granulocyte-monocyte progenitors; MEP, megakaryocyte-erythroid progenitors. **b, c**, Data are geometric mean fluorescence intensity (MFI ± s.e.m.) values. **d–k**, Data are mean ± s.e.m. values with *P* values from unpaired two-tailed *t*-test (see Methods for details on statistics).

on mouse leukaemic cells² as well as DNA proximity analysis using fluorescence *in situ* hybridization (DNA FISH) in LSK cells showed that the *Myc* promoter and BENC are in close physical proximity to each other (Extended Data Fig. 3i–k), suggesting that *cis*-regulation is mediated by chromosomal looping in haematopoietic stem and progenitor cells (HSPCs). To provide genetic evidence for this hypothesis, we generated compound heterozygous mice carrying *Myc*^{Δ15–17} on one chromosome and the *Myc*^{ΔORF} null allele¹⁸ on the other. We found that *Myc*^{Δ15–17/ΔORF} bone marrow cells failed to form normal colonies in colony-forming assays, similar to homozygous *Myc*^{ΔORF/ΔORF} cells¹⁸ (Extended Data Fig. 4a, b). Moreover, compound *Myc*^{Δ15–17/ΔORF} mice displayed a highly similar phenotype to the homozygous *Myc*^{Δ15–17} mutants (Fig. 2a, b and Extended Data Fig. 4c), demonstrating allelism between the two distant genetic elements. Moreover, whereas *Myc* levels were reduced by less than twofold in *Myc*^{Δ15–17} or *Myc*^{ΔORF} heterozygous cells, its expression was strongly decreased or even completely lost in *Myc*^{Δ15–17/ΔORF} animals in HSC, MPP, common myeloid progenitors, granulocyte-monocyte progenitors, common lymphoid progenitor populations and in most mature cell types, except for T cells and megakaryocyte-erythroid progenitors (Fig. 2c). Of note, homozygous *Myc*^{Δ15–17} mice showed normal *Myc* expression in all non-haematopoietic tissues analysed (Extended Data Fig. 4d and ref. 5). Reduced *Myc* expression due to BENC deletion also led to upregulation of *Mycn* in many blood cell types, with the noticeable exception of HSCs (Extended Data Fig. 4e). Taken together, these data demonstrate that BENC is a distantly located but essential lineage-specific *cis*-acting enhancer regulating *Myc* expression in HSCs and their progeny.

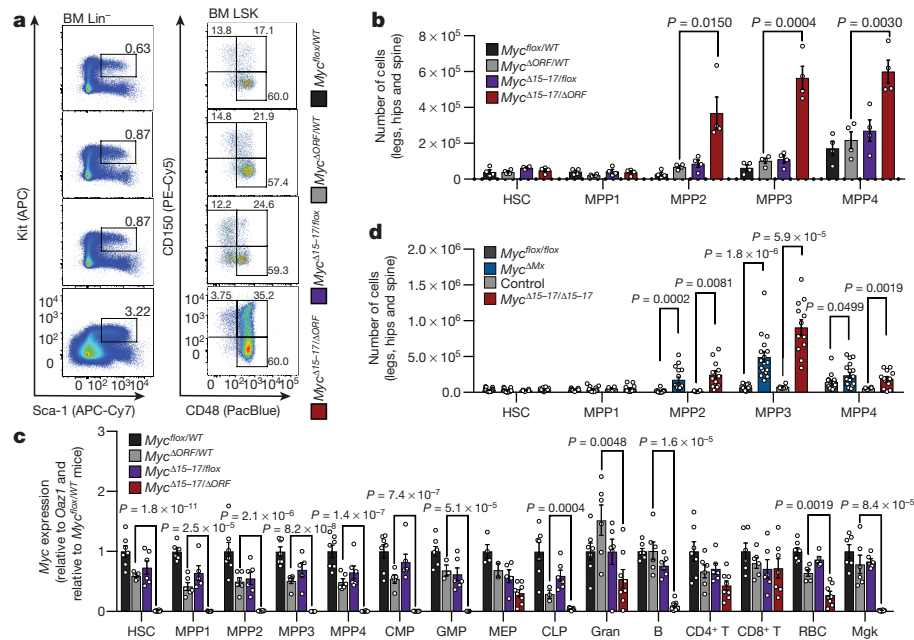


Figure 2 | BENC (15–17) directs *Myc* expression to haematopoietic cells and its deletion closely mimics *Mx-Cre*-mediated conditional deletion of the *Myc* gene. **a**, Representative expression profiles of Lin[−]-gated (left) and LSK-gated (right) bone marrow cells stained for the indicated cell surface markers by flow cytometry. **b**, Number of bone marrow HSCs and MPP1–4 cells. **c**, Relative *Myc* mRNA expression in the indicated

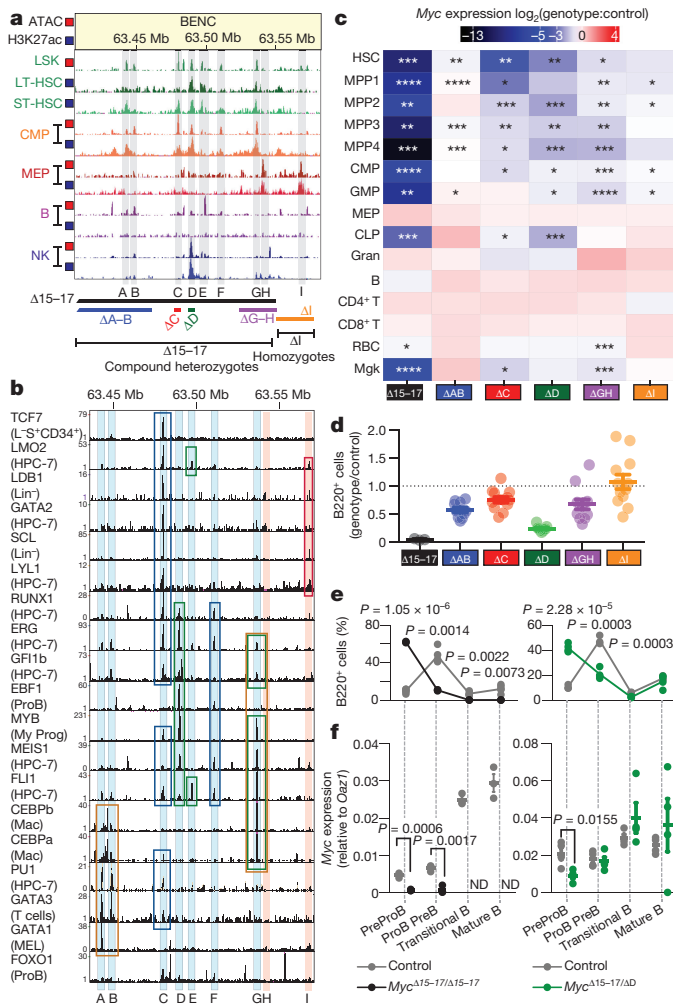
haematopoietic bone marrow cell populations were determined by reverse-transcription quantitative PCR (RT-qPCR). **d**, Number of HSCs and MPP1–4 cells in bone marrow of homozygous *Myc*^{Δ15-17/Δ15-17} mice and mice in which *Myc* was conditionally inactivated by *Mx-Cre* (*Myc*^{ΔMx}) plus controls. Mean ± s.e.m. values are shown and *P* values from unpaired two-tailed *t*-tests are reported (see Methods for details on statistics).

Although homozygous *Myc*^{Δ15-17} mice showed similar changes in the HSPCs, myeloid and B cell populations as our previous model of *Mx-Cre*-mediated inactivation of *Myc* in HSCs³, they did not develop the severe anaemia or the loss of thymocytes that had been observed in the latter model^{3,19}. Therefore, we repeated this analysis using the same experimental conditions as in the *Mx-Cre* model. Both models show a similar accumulation of MPP2–4 cells (Fig. 2d) and reduction in total bone marrow cellularity due to a decreased number of committed progenitors and loss of myeloid and B cells (Extended Data Fig. 4f–j). However, whereas T cell development in the thymus was completely abolished after conditional elimination of the *Myc* gene, *Myc*^{Δ15-17/Δ15-17} mice showed almost normal thymic T cell development with the exception of a transient increase in double-negative 3 thymocytes (Extended Data Fig. 4k–n). Similarly, the overall number of red blood cells hardly changed in *Myc*^{Δ15-17/Δ15-17} mutants, in contrast to the depletion of these cells after *Myc* deletion (Extended Data Fig. 4m, n). The minimal involvement of BENC in the differentiation of erythroid and T cell lineages is consistent with the moderately decreased or unchanged level of *Myc* expression in *Myc*^{Δ15-17/Δ15-17} megakaryocyte–erythroid progenitors and T cells (Fig. 2c), indicating that alternative regulatory regions are active in these lineages (Extended Data Fig. 5a for T cell development and refs 9, 10).

To gain further insight into the structure of BENC, we examined its chromatin composition in different blood cell types by chromatin immunoprecipitation followed by sequencing (ChIP-seq) and by assay for transposase-accessible chromatin using sequencing (ATAC-seq)²⁰ (Fig. 3a, b and Extended Data Fig. 5b–d). Several prominent peaks for the enhancer-associated mark H3K27ac were present in the different stem and progenitor cell types, whereas reduced enrichment of peaks was observed in mature cell types consistent with the activity mediated by BENC (Fig. 3a and Extended Data Fig. 5b). Using ATAC-seq and H3K27ac ChIP-seq data, we identified nine prominent modules, eight modules within *Myc*^{Δ15-17} (A–H) plus an adjacent module (I). All modules, with the exception of H, showed strong evolutionary conservation as analysed

by ‘genomic evolutionary rate profiling’ (Extended Data Fig. 5c, d). Many important haematopoietic regulators (for example, LYL1, SCL (also known as TAL1), FLI1, LMO2, RUNX1, MYB and MEIS1) that are recruited to different BENC modules do not appear to bind the *Myc* promoter, emphasizing that BENC constitutes the regulatory platform through which the combined inputs of these regulators are integrated (Fig. 3b and Extended Data Fig. 5d). H3K27ac enrichment in individual modules varied across different cell types, as did the recruitment of haematopoietic transcription factors (Fig. 3a, b and Extended Data Fig. 5d). Remarkably, broadly expressed transcription factors, such as GATA factors or PU.1, were recruited to different BENC modules in a cell-type-specific manner (Extended Data Fig. 5e, f), indicating that their binding may depend on cooperative interactions and/or on epigenetic priming of the corresponding modules. These tissue-specific profiles of the different modules suggest that cells may combine the activity of different sets of BENC enhancer modules in a combinatorial manner to implement *Myc* regulation in different cell types.

To functionally dissect the role of each module, we generated five mouse lines carrying deletions of individual (C, D and I) and adjacent modules (A–B and G–H) (Fig. 3a, bottom). Compared to the very strong effects observed in the *Myc*^{Δ15-17/Δ15-17} mutant (A–H deletion), deletions of one or two modules showed overall milder effects, and affected *Myc* expression in a more restricted manner that was cell-type-specific (Fig. 3c and Extended Data Fig. 6a). Deletion of modules C and D mainly affected *Myc* expression in HSCs and MPP1 cells, in agreement with strong H3K27ac enrichment in long-term HSCs (LSK CD135[−]CD34[−]) and short-term HSCs (LSK CD135[−]CD34⁺) in these modules (Fig. 3a, c and Extended Data Figs 5b, d, 6c). Deletion of BENC led to a strong decrease in B cell numbers (Fig. 3d and Extended Data Fig. 6b), indicating that it is required for B cell development. Closer examination of B cell development in mice that had individual modules deleted revealed a diversity of phenotypes. Module D is the only enhancer module for which a deletion showed a strong and significant reduction in B220⁺ B cells (Fig. 3d). This loss of B cells was associated with an accumulation of PreProB cells (B220⁺CD24[−]CD43⁺) and decreased



Myc expression (Fig. 3e, f and Extended Data Fig. 6d–f). Deletion of module C caused the same effect on PreProB cells, although B220⁺ cells did not decrease as strongly (Extended Data Fig. 6g, h). By contrast, module A–B deletion resulted in increased *Myc* expression in mature B cells (B220⁺CD24⁺CD43⁺IgD⁺) (Extended Data Fig. 6g, h). Overall, these data and ATAC-seq data from B cell developmental stages (Extended Data Fig. 6i) indicate that BENC acts in a multi-modular manner: its activity results from variable combinations of independently acting enhancer modules, leading to cell-type-specific regulation of *Myc* expression levels.

BENC was initially identified as a potential *Myc* enhancer in a mouse model of acute myeloid leukaemia (AML) on the basis of chromatin features² (Extended Data Fig. 7a). Therefore, we investigated its importance for *Myc* expression in established leukaemia. For this, we first retrovirally transduced LSK cells from either *Mx-Cre;Myc*^{Δ15-17/flox} or *Mx-Cre;Myc*^{WT/flox} control mice with a construct expressing the *MLL-AF9* fusion gene. Transplantation of these LSK cells in sub-lethally irradiated recipient mice caused leukaemia. We transplanted the same number of transduced cells from these primary leukaemic mice into secondary recipient mice to obtain a cohort of leukaemic mice with highly reproducible disease onset. Eight days after transplantation, we injected mice with polyinosinic:polycytidylic acid (poly(I:C)) to induce Cre expression and deletion of the floxed *Myc* allele, thereby generating cells that retained only one functional *Myc* allele, with or without a copy of BENC in *cis* in cells from *Mx-Cre;Myc*^{WT/flox} (*Myc*^{WT/ΔMx}) or *Mx-Cre;Myc*^{Δ15-17/flox} (*Myc*^{Δ15-17/ΔMx}) donors (Fig. 4a). In mice transplanted with *Mx-Cre;Myc*^{WT/flox} leukaemic control cells, we observed a rapid accumulation of leukaemic cells in the peripheral blood, even after consecutive poly(I:C) treatments (Fig. 4b), which resulted in the

death of recipient mice within 30–40 days (Fig. 4c). By contrast, in poly(I:C)-injected animals transplanted with *Mx-Cre;Myc*^{Δ15-17/flox} leukaemic cells, the disease accumulated much more slowly (Fig. 4b and Extended Data Fig. 7b). Simultaneously, myeloid differentiation markers (Gr1 and Mac1) were upregulated in *Myc*^{Δ15-17/ΔMx} leukaemic cells, suggesting that differentiation is induced after *Myc* expression is lost (Extended Data Fig. 7d–f). When treated with a single series of poly(I:C) injections, the onset of leukaemia was delayed; however, ultimately the leukaemia relapsed and the mice died (Extended Data Fig. 7c). Recurrent injection of poly(I:C), on the other hand, led to significantly prolonged survival (Fig. 4c), indicating that the relapse was mainly due to some residual leukaemic cells that failed to delete the floxed *Myc* allele after each round of Cre induction. Taken together, these data demonstrate that leukaemic cells are addicted²¹ to the presence of BENC, and suggest that a direct impairment of BENC activity may be sufficient to account for the effects of bromodomain inhibition that have been observed previously².

The position, sequence and architecture of BENC are conserved between mice and humans. ATAC-seq data show that the BENC modules are predominantly located within open chromatin in human HSCs, MPPs and committed progenitors, but not in most mature haematopoietic cell types, thus showing a similar activity pattern as in the mouse (Extended Data Fig. 8a, b). Promoter capture high-resolution chromosome conformation capture analysis and DNA FISH showed strong physical proximity between the *MYC* promoter and BENC in human CD34⁺ cells¹⁴ (Extended Data Fig. 8c). Supporting a role for BENC in human haematopoiesis, common single-nucleotide polymorphisms that are present within BENC have been associated with different human haematological traits, including monocyte counts^{22,23},

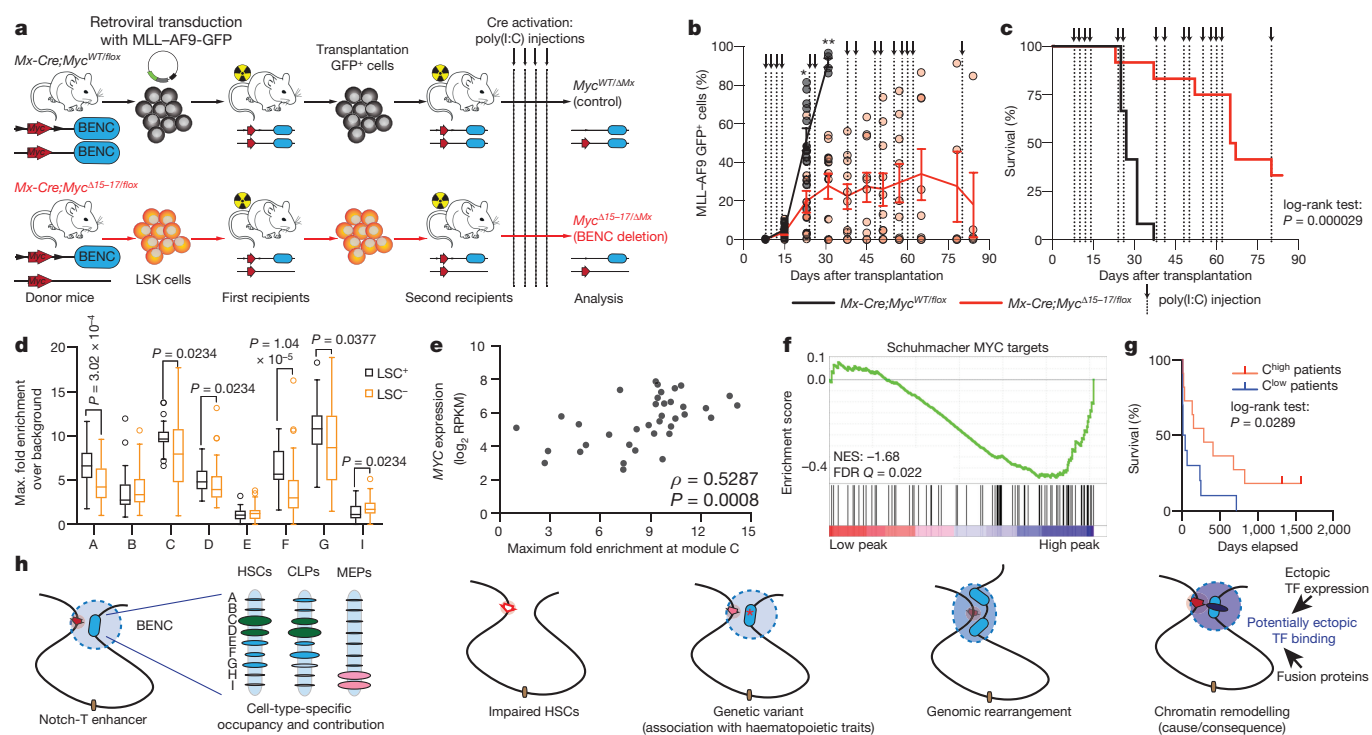


Figure 4 | BENC is required for maintenance of MLL-AF9 leukaemia in mice and accessibility to module C in human LSCs is linked to patient survival. **a**, Outline of the MLL-AF9 leukaemia models generated to mimic the poly(I:C)-induced acute loss of BENC in established leukaemia. **b**, **c**, Acute loss of BENC induced by consecutive poly(I:C) injections (dotted lines) leads to a significant reduction in leukaemic cells in the peripheral blood (**b**, data are mean \pm s.e.m., *P* values from unpaired two-tailed *t*-test, $*P = 7 \times 10^{-4}$, $**P = 5.96 \times 10^{-5}$) and is associated with significantly prolonged survival compared to mice in which BENC was retained (**c**, Mantel-Cox test). **d**, BENC enhancer module accessibility in human LSC⁺ and LSC⁻ populations as measured by ATAC-seq. The middle line indicates the median and the edges of the box and whiskers correspond

to the 75th and 25th percentile (box) plus and minus $1.5 \times$ inter-quartile range (whiskers), respectively; *P* values from two-tailed Wilcoxon rank-sum test. **e**, Correlation of the maximum ATAC-seq signal in module C with MYC expression (reads per kilobase per million reads (RPKM)) in primary samples using Spearman's rank correlation. **f**, Gene set enrichment of MYC target gene signature in module C^{high} LSC⁺ samples. FDR, false-discovery rate; NES, normalized enrichment score. **g**, Kaplan-Meier representation of patient overall survival according to ATAC-seq signal in module C in LSC⁺ fractions. **h**, Model showing BENC structure and function in normal haematopoiesis, after deletion, with genetic variants, amplifications (that is, leukaemia) and alterations in its activity by chromatin remodelling. TF, transcription factor. See Methods for details on statistics.

and recurrent focal chromosomal amplifications overlapping the C-F modules were observed in patients with AML^{24,25} (Extended Data Fig. 8a).

ATAC-seq performed on 93 functionally defined human AML leukaemic stem cell (LSC⁺) and blast populations (LSC⁻)^{26,27} isolated from primary patient samples revealed that module C shows a significantly greater accessibility in LSC⁺ compared to LSC⁻ fractions (Fig. 4d and Extended Data Fig. 8d). This differential accessibility is consistent with the role of module C in mouse HSCs, as the gene expression program of HSCs is highly similar to LSCs^{26,27}. Moreover, the ATAC-seq peak height of module C correlates with MYC expression, suggesting that MYC is regulated by module C in primary human leukaemic cells (Fig. 4e). In addition, samples from patients with AML with a high ATAC-seq peak in module C (*C*^{high}) showed a positive correlation with HSPC transcriptional programs and an anti-correlation with transcriptional programs of more differentiated cell types (Extended Data Fig. 8e). Furthermore, module C^{high} LSC⁺ fractions were enriched for mitotic MYC target genes²⁸ (Fig. 4f). In module C^{high} LSC⁻ fractions, we found several additional MYC gene sets that were enriched, which broadly cover genes regulated by MYC (Extended Data Fig. 8f). Taken together, these data indicate that *C*^{high} cells have higher MYC activity and increased proliferation. Notably, stratification of patients according to peak height in LSC⁺ fractions showed that patients with *C*^{high} have a significantly better overall survival, which was not the case for any of the other modules showing differential accessibility (Fig. 4g and Extended Data Fig. 8g). Importantly, this stratification is not correlated with white blood cell counts or bone marrow blast

counts (Extended Data Fig. 8h). A similar trend is observed in an independent AML cohort that consisted of immunophenotypically defined pre-leukaemic HSCs, LSCs and blasts²⁹ (Extended Data Fig. 8i-k). We suggest that the highly proliferative *C*^{high} LSCs may be more susceptible to anti-proliferative chemotherapy, resulting in a more favourable overall survival of *C*^{high} patients²⁷.

Here we show that *Myc* expression in the most haematopoietic lineages is to a large extent determined by a single cluster of enhancers (BENC). The combinatorial usage of enhancers is known to provide greater precision for the regulation of tissue-specific gene expression^{30,31}. Our data suggest that through the combined action of their constitutive modules, enhancer clusters, such as BENC (or the β -globin locus control region³²), can also yield variable outputs in cells that express related transcription factors. This property, which may be preferred by the adjacent organization of the enhancer modules, may have an important role in a cellular differentiation hierarchy, as fine-tuned expression of a given gene is important for the modulation of cell fate and phenotypes from stem cells to terminally differentiated cell types (Fig. 4h).

Such genomic organizations, in which adjacent and partly redundant enhancer modules recruit a diversity of transcription factors that act in a combinatorial manner, also provide many possibilities for the evolution of gene regulation. Genetic variations that alter the recruitment of transcription factors within one module may not interfere with the overall activity of the enhancer cluster, but can sensitize the cluster to variation in the expression level of a given transcription factor, enabling the modulation of its activity in response to cellular diversity. As such,

those regions may be primed to allow, over the course of evolution, the establishment of increasingly complex stem-cell-driven tissue hierarchies. From this aspect, these regions can be seen as ‘super-enhancers’, even though in a given cell type, their mode of action may not be very different from the activity of a collection of dispersed enhancer modules.

The conserved role of BENC in leukaemic cells underscores the importance of this region for the control of haematopoietic malignancies. The specific correlation between chromatin activity of individual BENC sub-enhancers, LSC biology and therapy responsiveness in patients with AML raises the possibility to develop epigenetic biomarkers, which have prognostic and predictive power, that could easily be assayed by ATAC-seq analysis of small amounts of DNA obtained from patients with leukaemia. Furthermore, our data suggest that BENC enhancer modules may be specific therapeutic targets for the treatment of *MYC*-addicted AMLs²¹.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 February 2015; accepted 4 December 2017.

Published online 17 January 2018.

- Delgado, M. D. & León, J. Myc roles in hematopoiesis and leukemia. *Genes Cancer* **1**, 605–616 (2010).
- Shi, J. *et al.* Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated *Myc* regulation. *Genes Dev.* **27**, 2648–2662 (2013).
- Wilson, A. *et al.* c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev.* **18**, 2747–2763 (2004).
- Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Uslu, V. V. *et al.* Long-range enhancers regulating *Myc* expression are required for normal facial morphogenesis. *Nat. Genet.* **46**, 753–758 (2014).
- Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
- The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Ahmadiyeh, N. *et al.* 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with *MYC*. *Proc. Natl Acad. Sci. USA* **107**, 9742–9746 (2010).
- Herranz, D. *et al.* A NOTCH1-driven *MYC* enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat. Med.* **20**, 1130–1137 (2014).
- Yashiro-Ohtani, Y. *et al.* Long-range enhancer activity determines *Myc* sensitivity to Notch inhibitors in T cell leukemia. *Proc. Natl Acad. Sci. USA* **111**, E4946–E4953 (2014).
- Sur, I. K. *et al.* Mice lacking a *Myc* enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* **338**, 1360–1363 (2012).
- Zhang, X. *et al.* Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat. Genet.* **48**, 176–182 (2016).
- Kieffer-Kwon, K. R. *et al.* Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* **155**, 1507–1520 (2013).
- Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
- Rathert, P. *et al.* Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. *Nature* **525**, 543–547 (2015).
- Fulco, C. P. *et al.* Systematic mapping of functional enhancer–promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
- Cabezas-Wallscheid, N. *et al.* Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell* **15**, 507–522 (2014).
- Trumpp, A. *et al.* c-Myc regulates mammalian body size by controlling cell number but not cell size. *Nature* **414**, 768–773 (2001).
- Guo, Y. *et al.* c-Myc-mediated control of cell fate in megakaryocyte–erythrocyte progenitors. *Blood* **114**, 2097–2106 (2009).

- Lara-Astiaso, D. *et al.* Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).
- Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional addiction in cancer. *Cell* **168**, 629–643 (2017).
- Nalls, M. A. *et al.* Multiple loci are associated with white blood cell phenotypes. *PLoS Genet.* **7**, e1002113 (2011).
- Okada, Y. *et al.* Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. *PLoS Genet.* **7**, e1002067 (2011).
- Kühn, M. W. M. *et al.* High-resolution genomic profiling of adult and pediatric core-binding factor acute myeloid leukemia reveals new recurrent genomic alterations. *Blood* **119**, e67–e75 (2012).
- Radtke, I. *et al.* Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia. *Proc. Natl Acad. Sci. USA* **106**, 12944–12949 (2009).
- Eppert, K. *et al.* Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat. Med.* **17**, 1086–1093 (2011).
- Ng, S. W. K. *et al.* A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**, 433–437 (2016).
- Schuhmacher, M. *et al.* The transcriptional program of a human B cell line in response to *Myc*. *Nucleic Acids Res.* **29**, 397–406 (2001).
- Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
- Montavon, T. *et al.* A regulatory archipelago controls *Hox* genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
- Marinić, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the *Fgf8* regulatory landscape. *Dev. Cell* **24**, 530–542 (2013).
- Li, Q., Peterson, K. R., Fang, X. & Stamatoyanopoulos, G. Locus control regions. *Blood* **100**, 3077–3086 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Spitz and Trumpp laboratories and colleagues for sharing reagents and helpful comments; A. Przybylla, M. Sohn and M. Neubauer for technical assistance; L. Alfaro and W. Zhang for sample preparation for RNA sequencing; M. Milsom for critical reading of the manuscript; R. Grosschedl and S. Boller for help with B cell development analysis; and the DKFZ Flow Cytometry Core facility and the EMBL and DKFZ Laboratory Animal Resource Facilities. Support was provided by PhD fellowships to V.V.U. (Jeff Schell Darwin Trust); M.P. (EMBL international PhD program), C.B. (Helmholtz International Graduate School for Cancer Research) and post-doctoral fellowship to S.R. (EMBL (EIPOD) under Marie Curie Actions COFUND). The J.E.D. and M.L. laboratories were supported in part by the Medicine by Design program (Toronto University), the Ontario Institute for Cancer Research, Cancer Stem Cell Consortium (OGI-047), the Canadian Institutes of Health Research and the CIHR-Japan Epigenetics in Stem Cells Program, Canadian Cancer Society, Terry Fox Foundation, and a Canada Research Chair to J.E.D. The A.T. laboratory was supported by the SFB 873 and FOR 2674 (Deutsche Forschungsgemeinschaft), the SyTASC consortium (Deutsche Krebshilfe) and the Dietmar Hopp Foundation.

Author Contributions C.B., L.v.P., V.V.U., A.T. and F.S. designed experiments. C.B., L.v.P., V.V.U., S.R., M.P., K.L., R.S., P.Z. and A.S.B. performed experiments related to mouse data. I.A. contributed to ChIP-seq analysis. A.M. and S.W.N. analysed RNA sequencing data, ATAC-seq and microarray data of patients with AML with conceptual input from C.B., M.L., J.E.D. and A.T. N.T. performed xenograft experiments. P.W.Z., M.L. and J.E.D. led and supervised the human studies. C.B., L.v.P., V.V.U., A.T. and F.S. analysed all other data. A.T. supervised and coordinated all haematopoiesis and leukaemia work as well as gene expression analyses. F.S. supervised the genetic strategies and gene regulation analyses. C.B., L.v.P., F.S. and A.T. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.T. (a.trumpp@dkfz.de) and F.S. (francois.spitz@pasteur.fr).

Reviewer Information *Nature* thanks B. Amati and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Mice. *Myc* alleles, insertions and some deletion strains have been described elsewhere^{3,5,18}. *Myc*^{ΔA-B}, *Myc*^{ΔG-H} and *Myc*^{ΔI} deletions were generated by TAMERE⁴³ (C57BL/6 background) using insertions obtained in the *Myc* locus⁶. *Myc*^{ΔC} and *Myc*^{ΔD} mice (mixed background) were produced by intra-cytoplasmic injection of *in vitro* transcribed human-codon-optimized *Cas9* (*hCas9*) mRNA (100 ng μl⁻¹) (from ref. 33) with two sgRNAs (50 ng μl⁻¹ each) flanking the respective enhancer modules in fertilized zygotes³⁴ (Supplementary Table 1). Deletion of the targeted module was confirmed by PCR amplification and sequencing. Targeted regions are reported in Supplementary Table 1. Compound heterozygous *Myc*^{Δ15-17/ΔORF} and control mice (*Myc*^{flox/WT}, *Myc*^{ΔORF/WT}, *Myc*^{Δ15-17/flox}) were obtained by crossing heterozygous *Myc*^{Δ15-17} mice (C57BL/6 background) with *Myc*^{flox/ΔORF} mice (mixed background). *Mx-Cre;Myc*^{WT/flox} and *Mx-Cre;Myc*^{Δ15-17/flox} mice were obtained by crossing heterozygous *Myc*^{Δ15-17} mice with *Mx-Cre;Myc*^{flox/flox} mice (C57BL/6 background). Deletion of flox alleles was induced by poly(I:C) (InvivoGen) injections. In brief, mice were injected 5× every other day with 5 mg kg⁻¹ poly(I:C). Mice were genotyped by PCR performed on DNA extracted from tail biopsies^{5,18} (primers sequences are available on request). In all experiments mice were age-, sex- and strain-matched. Mouse experiments performed at the DKFZ were performed according to protocols approved by the German authorities, Regierungspräsidium Karlsruhe (Z110/02, DKFZ299, G149-15, G150-15) and mouse experiments conducted at the EMBL were in accordance with the principles and guidelines in place at EMBL, as defined and overseen by its Institutional Animal Care and Use Committee, with the European Convention 18/3/1986 and Directives 86/609/EEC and 2010/63/EU.

Isolation of bone marrow cells. Age-matched mice (average age, 7.39 weeks; range, from 7 weeks to 14 weeks) were killed by cervical dislocation and bones from hind legs, forelegs and/or vertebrae were isolated. Bone marrow was extracted by crushing the bones using mortar and pestle in RPMI and 2% FCS. The suspension was filtered through a 40-μm mesh and centrifuged for 5 min at 1,500 r.p.m. Viable cell numbers were determined using the Vi-CELL cell viability analyser (Beckman Coulter).

Lineage depletion. Lin⁻ bone marrow was prepared by staining with a cocktail of monoclonal antibodies against the following lineage markers: CD4, CD8a, B220, Gr-1, Ter119 and CD11b. Labelled cells were then removed by incubation with sheep anti-rat IgG-coated Dynabeads (Life Technologies).

Antibodies. All FACS antibodies are commercially available and have been tested for use with mouse cells. The following antibodies were used (listed as: # (index), antigen, label, clone, category number, company, lot numbers): #1, B220, PE-Cy5, RA3-6B2, 15-0452, eBioscience, E06146-350, E06148-1632, E06147-1634; #2, B220, PE-Cy7, RA3-6B2, 25-0452, eBioscience, E07569-390, E07569-1630, E07569-1633, E07569-1635, E07569-1634, E07569-1636, E07569-1631, E07571-1630, E07569-1636; #3, B220, APC-eFluor780, RA3-6B2, 47-0452, eBioscience, E10028-413, 4275636, E10028-1630, E10028-1638, E10028-1631; #4, B220, biotin, RA3-6B2, 13-0452, eBioscience, E02530-301, E02528-1630, E02531-1633; #5, CD4, FITC, GK1.5, 11-0041, eBioscience, E00080-1630; #6, CD4, PE-Cy7, GK1.5, 25-0041, eBioscience, E07501-1429, E07501-1630, E07501-1632, E07501-1634, E07501-1633, E07501-1631; #7, CD4, eFluor450, GK1.5, 48-0041, eBioscience, E13684-105, E13684-106; #8, CD4, biotin, GK1.5, 13-0041, eBioscience, E02358-1632, E02358-1028, E02359-1631, E02359-1028; #9, CD8a, PE-Cy5, 53-6.7, 53-6.7, eBioscience, E06083-263, E06085-1631, E06085-1630; #10, CD8a, PE-Cy7, 53-6.7, 25-0081, eBioscience, E07510-1633, E07510-1634, E07510-1635, E07510-1632, E07510-1630, E029672; #11, CD8a, biotin, 53-6.7, 13-0081, eBioscience, E02385-1075, E02386-1633, E02386-1075; #12, CD11b, AlexaFluor700, M1/70, 56-0112, eBioscience, E033252, E08957-1633, E08957-1632, E08957-1630, 4293845; #13, CD11b, PE-Cy7, M1/70, 25-0112, eBioscience, E07514-1632, E07514-1633, E07514-1630; #14, CD11b, biotin, M1/70, 13-0112, eBioscience, 7595, E02410-1630, E02410-1631, E02411-1630, E02408-1630, E033766; #15, CD16/32 (FcγRII/III), eFluor450, 93, 57-0161, eBioscience, E08494-1630; #16, CD24, BV421, M1/69, 562563, BD Pharmingen, 5141676; #17, CD25, PE, PC61.5, 102008, eBioscience; #18, CD34, AlexaFluor700, RAM34, 56-0341, eBioscience, E08980-1237, E08980-1631, E17475-102, E17475-103, E08980-1632, E08980-1630; #19, CD34, FITC, RAM34, 11-0341, eBioscience, E00265-1393, E00265-1631, E00265-1633, E00265-1634, 4310179, E00265-1632, E00265-1630, E00265-1634; #20, CD41, eFluor450, MWReg30, 48-0411-82, eBioscience, 4275031; #21, CD41, FITC, MWReg30, 553848, BD Pharmingen, 38591, 72620; #22, CD41, PE-Cy7, MWReg30, 25-0411-80, eBioscience, E07563-1633, E07563-1634, E07563-1632; #23, CD43, APC, S7, 560663, BD Pharmingen, 4261684; #24, CD44, eFluor450, IM7, 103020, BioLegend, B127475; #25, CD45, AlexaFluor700, 30-F11, 56-0451, eBioscience, E031610, E08988-1631, E08988-1630; #26, CD45.1, FITC, A20.1, 11-0453, eBioscience, E00313-1483, E029243, E00313-1483, E00313-1630, E00315-1084; #27, CD45.1, PE, A20.1, 12-0453, eBioscience, E01252-1630; #28, CD45.1, PE-Cy7, A20.1, 25-0453, eBioscience, E07571-1573, E07571-1631,

E07571-1634, E07571-1633, E07571-1630; #29, CD45.2, Pacific Blue, 104, 109820, BioLegend, B137801, B169087, B129181, B145497, B195294, B176123, B156563, B141107, B219445; #30, CD45.2, AlexaFluor700, 104, 56-0454, eBioscience, E08994-1630, E08994-1631; #31, CD48, Pacific Blue, HM48-1, 103418, BioLegend, B127915, B170052, B218789, B143582; #32, CD48, PE, HM48-1, 12-0481, eBioscience, E01275-1634; #33, CD48, PE-Cy7, HM48-1, 103424, BioLegend, B123851, B193230, B131052, B141239; #34, CD71, PE, R17217, 12-0711, eBioscience, E01341-1632, E01341-1634; #35, CD117 (Kit), APC-eFluor780, 2B8, 47-1171, eBioscience, E08461-1350, E08461-1633, E08461-1634, E08461-1635, E08461-1632, E08461-1635; #36, CD127 (IL7Ra), PE, A7R34, 12-1271, eBioscience, E01474-1630, E01471-1630, E026170; #37, CD135, PE, A2F10, 12-1351, eBioscience, E01495-133, E01495-1633, E01495-1634, E01495-1632; #38, CD150, PE-Cy5, TC15-12F12.2, 115912, BioLegend, B135107, B150422, B150422, B124920, B164578, B197791, B164578, B142148; #39, Ki67, FITC, B56, 51-36524X, BD Pharmingen, 3213825, 88789; #40, Gr-1 (Ly6-G), APC, RB6-8C5, 17-5931, eBioscience, E07334-1630, E07334-1633, E07334-1631, E07334-1633, E07334-1632, E07334-340; #41, Gr-1 (Ly6-G), PE-Cy7, RB6-8C5, 25-5931, eBioscience, E07648-1631, E07648-1632, E07648-1633, E07648-1634, E07648-1630, 4295927; #42, Gr-1 (Ly6-G), biotin, RB6-8C5, 13-5931, eBioscience, E03074-1630, E03075-1393, E03072-1630, E033863; #43, IgD, BV510, 11-26c.2a, 563110, BD Pharmingen, 6049758; #44, IgM, FITC, R6-60.2, 553408, BD Pharmingen, 5215530; #45, Sca-1, APC, D7, 17-5981, eBioscience, E07354-217, E07354-217, E07355-1631, E07355-1632, E07355-1630, E07355-217, E07355-1630; #46, streptavidin, PE-TexasRed, not applicable, 551487, BD Pharmingen, 53741, 17281, 73206, 4085873; #47, Ter119, APC-eFluor780, TER-119, 47-5921, eBioscience, E10236-101, E10236-1630, E10236-1638, E10236-1639, E10236-1632, E10236-1631; #48, Ter119, PE-Cy7, TER-119, 25-5921, eBioscience, E07646-485, E07646-1631, E07646-1634, E07646-1632, E07646-1630; #49, Ter119, biotin, TER-119, 13-5921, eBioscience, E03070-1238, E03071-1632, E03068-1630, E03071-1630.

Flow cytometry and cell sorting. Populations of interest and corresponding antigens are listed in Extended Data Table 2. Antibodies were diluted in a 1:1 mix of 2.4G2 blocking buffer and RPMI and 2% FCS except for staining solutions that contained FcγRII/III-binding antibodies. Absolute cell numbers were calculated by multiplying FACS-determined frequencies with viable bone marrow cell numbers. FACS-Gal analysis was performed on bone marrow cells intracellularly stained with FDG (Sigma-Aldrich) according to the protocol from ref. 35. Samples were analysed on a BD LSRII flow cytometer or a BD LSRFortessa cell analyser. Cell sorting and sample analysis was executed on a BD FACSARIA, FACSARIAII, FACSARIAIII or FACSARIA Fusion. FACS data were analysed with FlowJo software v9-v10 (FlowJo LLC). Cell numbers were calculated using the frequency of total cells, as determined by FlowJo, and the total bone marrow cellularity (either legs, hips and spine preparation or just legs and hips). Gating strategies are shown in Supplementary Fig. 1.

Retroviral bone marrow transduction assays and MLL-*AF9* mouse model. LSK cells were isolated from bone marrow of *Mx-Cre;Myc*^{WT/flox} and *Mx-Cre;Myc*^{Δ15-17/flox} mice and transduced with MLL-*AF9*-IRES-GFP-encoding retroviruses³⁶. M. Milsom provided plasmids. Transduced cells were transplanted into sub-lethally irradiated recipient mice (5 Gy). AML progression was measured through facial vein bleed and FACS analysis. Bone marrow cells and splenocytes of terminally ill mice were isolated and 100,000 MLL-*AF9* GFP⁺ cells were transplanted into sub-lethally irradiated C57BL/6 Ly5.1 recipient mice. Mice were injected with 5 mg kg⁻¹ poly(I:C) to induce Cre expression from the Mx1 promoter. Disease progression was monitored by facial vein bleed and mice were euthanized when showing signs of disease/distress. For *Myc* expression analysis of leukaemic cells after deletion, GFP⁺CD45.2⁺ leukaemia cells were sorted from peripheral blood after erythrocyte lysis (ACK lysing buffer, Lonza). Experiments were performed in accordance with protocols approved by the local authorities (Regierungspräsidium Karlsruhe; G150-15) with no specific restrictions to the maximum percentage of leukaemic cells in the peripheral blood (measurements are reported in the Source Data of Fig. 4b, Extended Data Fig. 7b and in Supplementary Table 2).

Bone marrow transplantations. For generation of competitive chimaeras (80% donor bone marrow and 20% competitor bone marrow), mice were lethally irradiated with a 10-Gy split-dose 4h apart and bone marrow transplantation was performed within 24h by tail vein injection. For analysis of the potential of *Myc*^{Δ15-17/Δ15-17} bone marrow to produce T cells after transplantation, NSG mice were irradiated with 1.75 Gy and transplanted with 20,000 LSK cells combined with 400,000 LS⁻K cells. Engraftment was measured monthly through facial vein bleed and FACS analysis.

Haematopoietic colony-forming cell assay. Fresh total bone marrow was counted and diluted to a concentration of 2 × 10⁵ cells per ml in sterile PBS. For duplicates, 300 μl of this suspension was added to 3 ml MethoCult GF M3434, containing recombinant mouse SCF, recombinant mouse IL-3, recombinant human IL-6, and recombinant human EPO (STEMCELL Technologies). After 10 days, the dishes were scored for haematopoietic colonies.

RNA isolation and RT-qPCR. For RNA isolation, 500–50,000 cells were sorted into Extraction Buffer (Arcturus PicoPure, Applied Biosystems and Thermo Fisher Scientific) and RNA isolation was performed according to the manufacturer's instructions including RNase-free DNase digestion (Qiagen). RNA samples were transcribed into complementary DNA (cDNA) using the SuperScript VILO cDNA synthesis kit according to the manufacturer's instructions (Invitrogen). RT-qPCR measurement of individual cDNAs was performed using ABI Power SYBR Green Master Mix (Life Technologies). RT-qPCR reactions were carried out on a ViiA7 (Applied Biosystems) with the following primers: *Oaz1* forward, 5'-TTTCAGCTAGCATCCTGTACTCC-3'; *Oaz1* reverse, 5'-GACCTGGTCTTGTCTGTTAGA-3'; *Myc* forward, 5'-CCCTAGTGTGCTGATGAGGAGACAC-3'; *Myc* reverse, 5'-CCACAGACACCACA TCAATTTCTTCC-3'; *Mycn* forward, 5'-CTCCGAGAGGATACCTTGA-3'; *Mycn* reverse, 5'-TCTCTACGGTGACCACATCG-3'; *Gusb* forward, 5'-CTCTGGT GGCCTTACCTGAT-3'; *Gusb* reverse, 5'-CAGTTGTTGTACCTTCACCTC-3'; *Cd45* forward, 5'-GAGCAGACCCGAGATCCAC-3'; *Cd45* reverse, 5'-GCAGCACT ACCAGAAAAGGCA-3'; *Actb* forward, 5'-CTAAGGCCAACCGTGAAG-3'; *Actb* reverse, 5'-ACCAAGGCATACAGGGACA-3'.

DNA FISH. BAC-containing bacterial strains were purchased from Chori. Purified BACs were verified by PCR. RP23-397P6 was verified by primer 1944F: 5'-TCACTGTGTGTCATTGGCATA-3' and primer 1945R: 5'-GTGTCC ACTGGGAAGAGGA-3'. RP23-22L9 was verified by primer 1954F: 5'-ACTCC TTGCTCCCTTCTTC-3' and primer 1955R: 5'-CAGTGGGGCAATTGAG TCTT-3'. RP23-207P4 was verified by primer 1958F: 5'-GATGCCTACTTTGCC CTCAG-3' and primer 1959R: 5'-GGCGTCTGCCTGGTACTTACT-3'. BACs were labelled using Kretech ULS Products, FISHBright 495-Green (FLK-002) and FISHBright 550-Red (FLK-004) according to the manual.

LSK cells from wild-type C57BL/6 animals (8–12 weeks old) were initially sorted and fixed in a cold methanol:acetic acid (3:1) solution for 10 min. They were then treated with 0.075 M KCl for 5 min at 37°C and washed once in 2× SSC (0.3 M NaCl, 30 mM trisodiumcitrate, pH 7.20) at 37°C. The cells were dehydrated in 70%, 90% and 100% ethanol solutions, for 3 min each at room temperature. Then the cells were plated on poly-L-lysine-coated slides, air dried for 30 min at room temperature and the slides were gradually frozen and kept at −20°C until use. Right before use the slides were cooked at 42°C for 1 h to increase the attachment of the cells to the slide. The cells were dehydrated again once in 80%, once in 95% and twice in 100% ethanol at room temperature, 5 min each. Rehydration was performed by washing twice in 2× SSC at room temperature for 5 min. Subsequently, the cells were incubated in RNase mix (2 µl RNase H, 10 µl RNase A in 1 ml 2× SSC) at 37°C for 1 h. RNase mix was washed off by washing twice in 2× SSC and the same dehydration series (80%, 95% and twice 100% ethanol) was carried out. The slides were completely dried. After that, they were placed in formaldehyde solution (50% formaldehyde in 2× SSC) in a humid chamber at 80°C for 37 min. The slides were then washed three times in ice-cold 2× SSC. Then, 4 µl of BAC probes was mixed with hybridization buffer (2× SSC, 20% dextran sulphate, 10× BSA, 10% ribonucleoside vanadyl complex) in a total volume of 20 µl. BAC probes in hybridization buffer were denatured at 75°C or 10 min and immediately applied on the cells. Hybridization was carried out at 42°C in a dark chamber humidified with formaldehyde solution for 3 h to overnight. The probes were washed off by three consecutive washes with formaldehyde solution at 42°C, 7 min each and three washes with 2× SSC solution at 42°C, 5 min each. The cells were mounted with Vectashield DAPI (H-1200) and sealed.

PerkinElmer Imposition Ultraviolet Voxel Spinning Disk Confocal was used to visualize the DNA FISH slides. A 100×/1.3 oil objective was used to capture the images. Volocity 6.3 was used as an imaging and analysis software. Owing to the very low cell density, each slide was visualized at once and the position of each cell was marked on the coordinate plane of the slide in Volocity to avoid any duplication of the data. The following four laser lines were used: 405 nm for DAPI, 488 nm for AlexaFluor495, 561 nm for AlexaFluor550 and 640 nm for background signal. For each cell the *z* axis was defined between the top and bottom of the AlexaFluor550 staining. Images along the *z* axis were captured in 150 nm steps.

The 3D image analysis was done in Volocity. The following function was written to detect the 3D structure of the BAC probe signals: 'find object' in a given channel with the 'coarse' or 'very coarse' filter by using the percentage of intensity for a given 'region of interest'. Three-dimensional objects were automatically selected and manually curated according to the Extended 3D View in Volocity. The coordinates of centre of mass for each 3D object were exported to an Excel sheet. The distance between the centres of mass were calculated according to the following formula:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

where *d* is the distance; *x*₁, *y*₁, *z*₁ are the coordinates of the centre of mass of a 3D object obtained by the signal of a BAC probe; *x*₂, *y*₂, *z*₂ are the coordinates of the

centre of mass of a 3D object obtained by the signal of the other BAC probe. To avoid potential artefacts (for example, spots from homologous chromosomes), we removed the few samples where distances were above 1 µm (inclusion of those samples will only reinforce the differences observed). Distances are shown as a box plot.

Two-dimensional images were imported from Volocity as TIFF files. Images of a single *z*-axis layer were extracted for each channel. They were smoothed by Gaussian Filter (radius:2). The signal for each channel was emphasized by subtracting the background signal and multiplying with a certain integer. Different channels were merged to obtain the final images. In order to avoid counting auto-fluorescence as a signal, largely overlapping 3D signals were compared to the far-red channel signal. The signal that overlaps with the signal in the far-red channel was eliminated manually from the analysis.

ATAC-seq. Single-end sequencing reads were aligned using BWA 0.7.2³⁷ against the hg19 reference genome. Duplicate reads were filtered, and any reads mapping to poorly alignable regions were removed⁷. MACS 2.1.1³⁸ (with signal per million reads correction) was used to call significantly enriched peaks and compute fold enrichment of signal versus local background.

RNA sequencing. RNA was extracted from bulk peripheral blood mononuclear cells using the RNeasy Micro kit (Qiagen). For each sample, approximately 10 ng of total RNA was processed using the SMART cDNA synthesis protocol including SMARTScribe Reverse Transcriptase (Clontech, 639536). This method uses a modified oligo(dT) primer to prime the first strand synthesis reaction and a template switching mechanism to generate full-length single-stranded cDNAs containing the complete 5' end of the mRNA as well as universal priming sequences for end-to-end amplification during 20 cycles of PCR. The amplified cDNA was subjected to automated Illumina paired-end library construction using the NEBNext paired-end DNA sample Prep Kit (NEB, E6000B-25). Libraries were sequenced on Illumina HiSeq2000 instruments with an average of approximately 161 million Chastity-passed paired reads of 75 bp in length per sample. Sequence data were aligned using the BWA software version 0.5.7 to the GRCh37-lite human reference genome. Reads per kilobase per million reads (RPKM)³⁹ values were calculated using the formula: (number of reads mapped to all exons in a gene × 1,000,000,000)/(NORM_TOTAL × sum of the lengths of all exons in the gene) where NORM_TOTAL is the total number of reads that are mapped to exons (that is, fractional read count for exons), excluding those belonging to the mitochondrial chromosome. These data were used to calculate the correlation between the ATAC-seq peak height in module C and MYC expression.

Survival analysis of patients with AML. Overall survival was defined as the time from AML diagnosis until death from any cause or last clinical follow-up. Univariate survival analysis was performed using Kaplan–Meier and Cox proportional hazards models with comparisons performed using log-rank tests. Wald's test was used to evaluate the significance of hazard ratios. All survival analyses were performed using the survival⁴⁰ v.2.38-1 R package).

Correlation analysis. All statistical analyses were performed in R v.3.1.0, and all data comparisons with a *P* value of less than 0.05 were statistically significantly different. The Spearman rank method of correlation was used unless specified otherwise. Because the relationship between ATAC-seq peak height and gene expression is not known, the Spearman rank method of correlation was used, which assumes a monotonic relationship between the two analysed variables. Pearson's product-moment correlation was used to analyse the relationship between clinical patient parameters and module C ATAC-seq peak height in LSC⁺ and LSC[−] fractions.

Microarray gene expression. RNA was extracted from 23 unsorted samples from patients with AML, 21 LSC⁺, and 19 LSC[−] fractions, which were then analysed by gene expression measurements using the Illumina human HT-12 v4 microarray platform. Gene expression was normalized using the lumi v.2.16.0 R package⁴¹. These data were used for analysis using the perturbation model (see 'Estimation of relative cell-type-specific composition of AML samples') and gene set enrichment analysis (see 'Gene set enrichment analysis') for Myc signatures.

Estimation of relative cell-type-specific composition of AML samples. Relative proportions of transcriptional programs of various blood cell types including HSC, MPP, CMP, ETP, MEP, monocytes, granulocytes, pro-B cells and other cell types phenotypically purified from human umbilical cord blood (Gene Expression Omnibus (GEO) datasets GSE42414⁴² and GSE24759⁴³) composing AML transcriptional profiles were assessed using the perturbation model (PERT)⁴⁴. These estimates of relative proportions of cell-type-specific transcriptional programs along with module C peaks in LSC⁺- and LSC[−]-sorted cell fractions were then used as input for the Spearman correlation analysis to determine the nature of their association.

Gene set enrichment analysis. Gene set enrichment analysis⁴⁵ v.2-2.2. was performed with 2,000 gene set permutations on LSC⁺ and LSC[−] cell fractions to determine whether MYC target gene sets were overrepresented in samples with above or below median module C peaks. Gene sets were obtained from refs 28, 46, 47.

Patient samples and xenograft experiments. All biological samples were collected with informed consent according to procedures approved by the Research Ethics Board of the University Health Network (UHN, REB 01-0573-C) and viably frozen in the Princess Margaret Leukaemia Bank. LSC⁺ and LSC⁻ status was scored using the methods as outlined in ref. 26. Patient parameters are reported in Supplementary Table 3.

Sample size considerations. For animal experiments, we aimed for at least three animals per group (range 3–17) to allow basic statistical inference while using a justifiable number of mutant mice.

Data exclusion. To avoid potential artefacts for DNA FISH experiments (for example, spots from homologous chromosomes), we removed a few samples in which distances were above 1 μ m (inclusion of those samples will only reinforce the differences observed). Otherwise, unless specifically mentioned, no data were excluded from the analyses.

Replication. All attempts at replication were successful.

Randomization. Allocation of mice to groups was not formally randomized. However, the possible confounders ‘experimenter’ and ‘day of experiment’ were equally matched between groups.

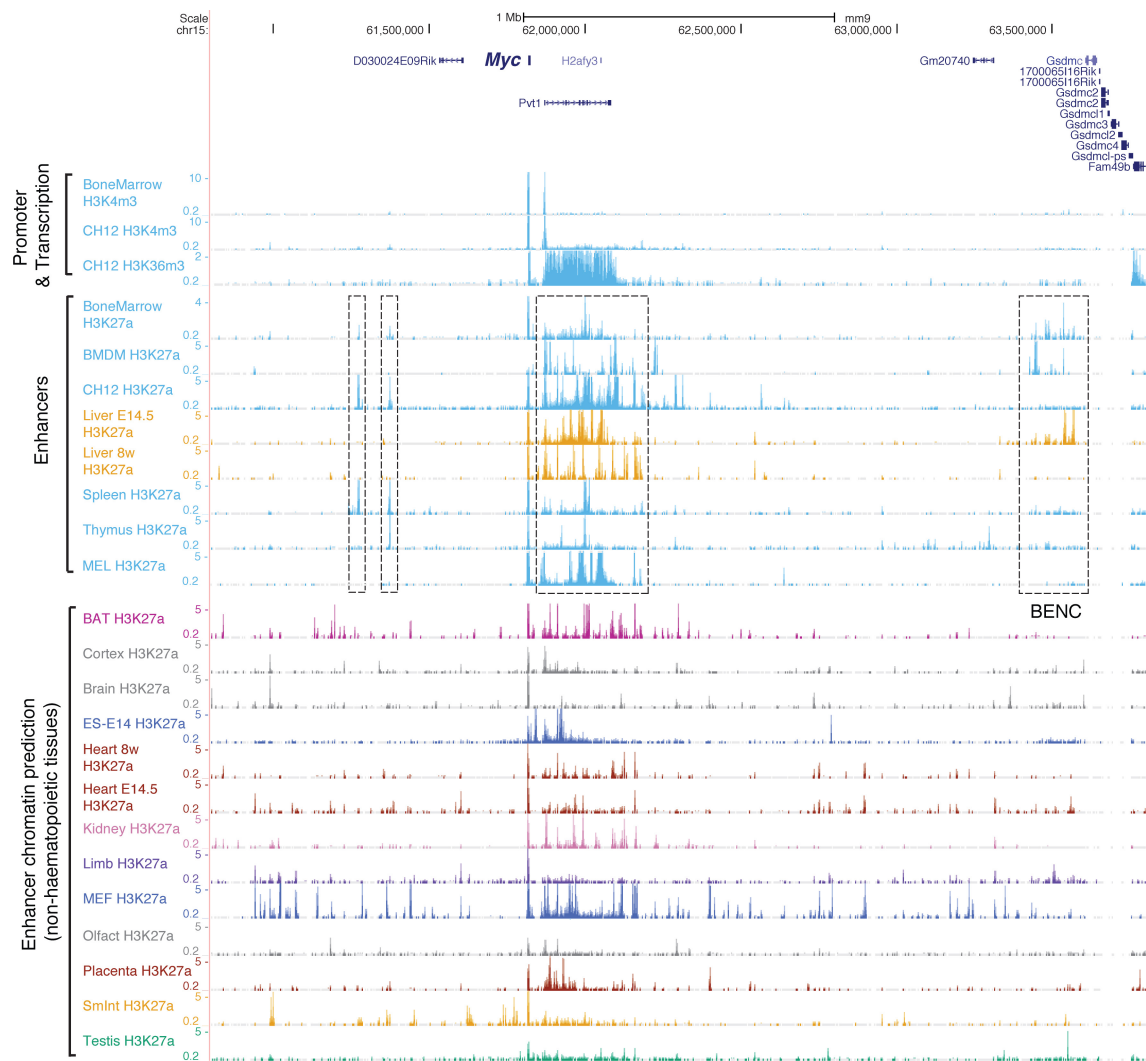
Blinding. No experiments were blinded.

Statistical analysis. Data were processed using Prism v.6 and v.7 (GraphPad Software). All analyses were performed using two-tailed Student’s *t* tests. If not otherwise indicated, the uncertainty in the mean is reported as the standard error of the mean (s.e.m.), which also takes the sample size into account. Sample sizes for each dataset within the Figures were as follows: Fig. 1b, *n* = 3 mice per group (data from two independent experiments); Fig. 1c, representative data from three independent experiments, *n* = 8 mice; Fig. 1d, *n* = 8 mice; Fig. 1e, *n* = 5 mice; Fig. 1f, *n* = 1 (input at 0 weeks), *n* = 6 mice (control, all other data points), *n* = 7 mice (*Myc* ^{Δ 15-17/ Δ 15-17}, all other data points); Fig. 1g, *n* = 6 mice (control), *n* = 7 mice (*Myc* ^{Δ 15-17/ Δ 15-17}); Fig. 2a, representative data from two independent experiments; Fig. 2b, *n* = 4 mice (all groups), one representative experiment from two independent experiments is shown; Fig. 2c, *n* = 17 mice (*Myc*^{flx/flx}), *n* = 17 mice (*Myc* ^{Δ Mx}), *n* = 6 mice (control), *n* = 12 mice (*Myc*^{WT/flx}), data from two independent experiments are shown; Fig. 2d, *Myc*^{WT/flx} (*n* = 7 mice for all populations, except MPP1 (*n* = 6 mice), MEP (*n* = 4 mice), CD8⁺ T (*n* = 6 mice), megakaryocytes (*n* = 6 mice)), *Myc* ^{Δ ORF/WT} (*n* = 6 mice for all populations, except HSC (*n* = 5 mice), CMP (*n* = 5 mice), GMP (*n* = 3 mice), MEP (*n* = 2 mice), CLP (*n* = 3 mice), RBC (*n* = 5 mice)), *Myc* ^{Δ 15-17/flx} (*n* = 6 mice for all populations, except MPP1 (*n* = 5 mice), MPP3 (*n* = 5 mice), MEP (*n* = 5 mice), RBC (*n* = 5 mice)), *Myc* ^{Δ 15-17/ Δ ORF} (*n* = 8 mice for all populations, except MPP1 (*n* = 7 mice), MPP3 (*n* = 7 mice), GMP (*n* = 5 mice), CLP (*n* = 7 mice), CD8⁺ T (*n* = 7 mice)), data from two independent experiments are shown; Fig. 3c, *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ 15-17} mice (*n* = 4 mice for all other populations, except MPP1, MPP3, MEP and RBC (*n* = 3 mice)), *Myc* ^{Δ 15-17/ Δ 15-17} mice (*n* = 4 mice for all populations), *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ A-B} mice (*n* = 5 mice for all populations, except RBC (*n* = 4 mice) and B cells (*n* = 3 mice)), *Myc* ^{Δ 15-17/ Δ A-B} mice (*n* = 5 mice for all populations, except RBC (*n* = 4 mice)), *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ C} mice (*n* = 5 for all populations, except HSCs, MPP1, MPP2, MPP3 and MPP4 (*n* = 4 mice)), *Myc* ^{Δ 15-17/ Δ C} mice (*n* = 6 mice for all populations, except CD4⁺ T and CD8⁺ T (*n* = 5 mice)), *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ D} mice (*n* = 4 mice for all populations), *Myc* ^{Δ 15-17/ Δ D} mice (*n* = 4 mice for all populations), *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ G-H} mice (*n* = 4 mice for all populations, except RBC (*n* = 3 mice)), *Myc* ^{Δ 15-17/ Δ G-H} mice (*n* = 5 mice for all populations, except HSC, MPP1, MPP2, MPP3 and MPP4 (*n* = 4 mice)), *Myc*^{WT/WT} control for *Myc* ^{Δ I/ Δ I} mice (*n* = 4 mice for all populations, except MEP (*n* = 3)), *Myc* ^{Δ I/ Δ I} mice (*n* = 4 mice for all populations, except CD8⁺ T (*n* = 3 mice)), data from one experiment are shown; Fig. 3d, *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ 15-17} (*n* = 3 mice), *Myc* ^{Δ 15-17/ Δ 15-17} mice (*n* = 3 mice), *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ A-B} (*n* = 10 mice), *Myc* ^{Δ 15-17/ Δ A-B}

(*n* = 10 mice), *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ C} (*n* = 14 mice), *Myc* ^{Δ 15-17/ Δ C} (*n* = 14 mice), *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ D} (*n* = 14 mice), *Myc* ^{Δ 15-17/ Δ D} (*n* = 9 mice), *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ G-H} (*n* = 8 mice), *Myc* ^{Δ 15-17/ Δ G-H} (*n* = 12 mice), *Myc*^{WT/WT} control for *Myc* ^{Δ I/ Δ I} (*n* = 11 mice), *Myc* ^{Δ I/ Δ I} (*n* = 12 mice), Fig. 3e, *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ 15-17} (*n* = 5 mice); *Myc* ^{Δ 15-17/ Δ 15-17} (*n* = 3 mice); *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ D} (*n* = 3 mice); *Myc* ^{Δ 15-17/ Δ D} (*n* = 4 mice), data from one experiment are shown; Fig. 3f, *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ 15-17} (*n* = 3 mice); *Myc* ^{Δ 15-17/ Δ 15-17} (*n* = 3 mice for PrePro B and Pro B Pre B and others not determined); *Myc* ^{Δ 15-17/WT} control for *Myc* ^{Δ 15-17/ Δ D} (*n* = 4 mice); *Myc* ^{Δ 15-17/ Δ D} (*n* = 4 mice), data from one experiment are shown; Fig. 4b, *Mx-Cre*; *Myc*^{WT/flx} (*n* = 12 mice for 8 days, 15 days and 23 days, *n* = 4 mice for 31 days), *Mx-Cre*; *Myc* ^{Δ 15-17/flx} (*n* = 12 mice for 8 days and 15 days, *n* = 11 mice for 23 and 31 days, *n* = 10 mice for 38, 45 and 51 days, *n* = 9 mice for 57 days, *n* = 8 mice for 65 days, *n* = 5 for 78 and 84 days), data from one experiment are shown; Fig. 4c, *n* = 12 mice per group, data from one experiment are shown; Fig. 4d, *n* = 41 LSC⁺ fractions and 52 LSC⁻ fractions; Fig. 4e, *n* = 37 fractions; Fig. 4f, 21 LSC⁺ fractions; Fig. 4g, 11 patients in the module C^{high} group and 10 patients in the module C^{low} group.

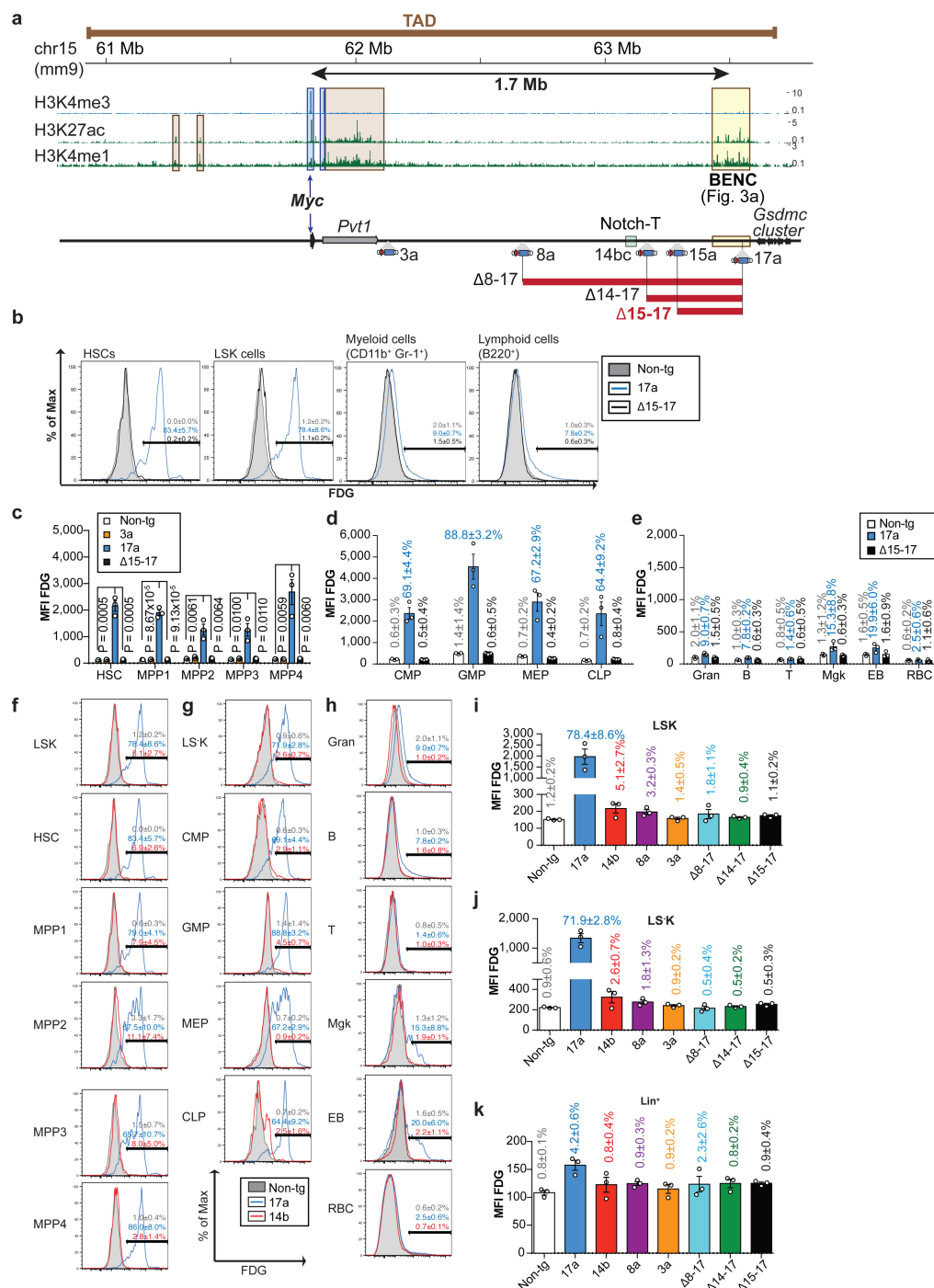
Data availability. Publically available datasets used in this study are reported in Extended Data Table 1 or directly mentioned with accession numbers in the text. Source Data for Figs 1–4 and Extended Data Figs 2–4, 6–8 are included in the online version of the paper. Sequencing and expression data related to LSC⁺ and LSC⁻ samples from patients with AML are available from J.E.D. (john.dick@uhnresearch.ca) upon reasonable request.

33. Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
34. Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
35. Guo, W. & Wu, H. Detection of LacZ expression by FACS-Gal analysis. *Protoc. Exch.* <https://doi.org/10.1038/nprot.2008.163> (2008).
36. Krivtsov, A. V. et al. Transformation from committed progenitor to leukaemia stem cell initiated by MLL–AF9. *Nature* **442**, 818–822 (2006).
37. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
38. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
39. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* **5**, 621–628 (2008).
40. Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model* (Springer, 2000).
41. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).
42. Laurenti, E. et al. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat. Immunol.* **14**, 756–763 (2013).
43. Novershtern, N. et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309 (2011).
44. Qiao, W. et al. PERT: a method for expression deconvolution of human blood samples from varied microenvironmental and developmental conditions. *PLOS Comput. Biol.* **8**, e1002838 (2012).
45. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
46. Liberzon, A. et al. The Molecular Signatures Database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
47. Kim, J. et al. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* **143**, 313–324 (2010).
48. Heng, T. S. P. & Painter, M. W. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).



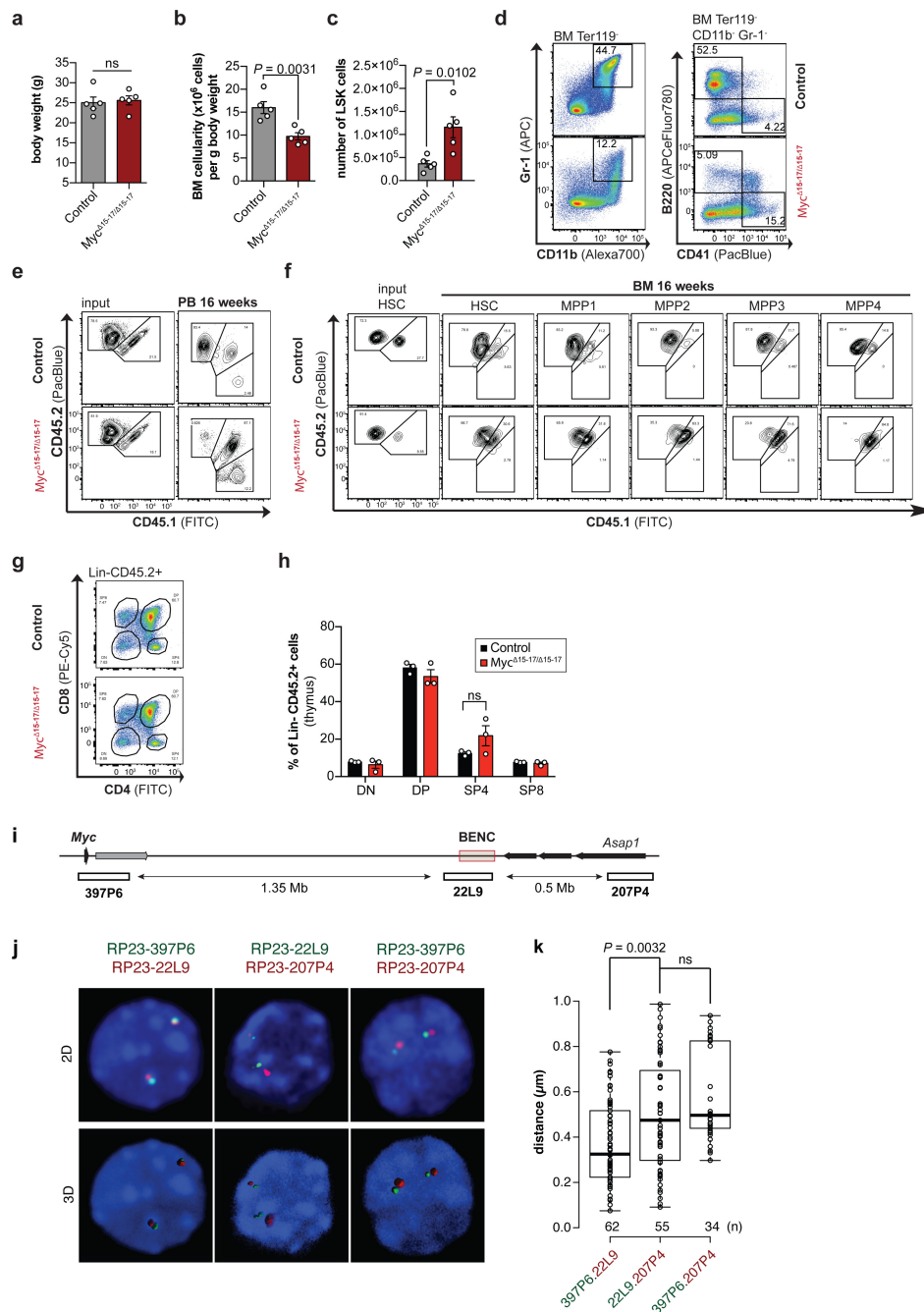
Extended Data Figure 1 | Chromatin profiles at the *Myc* locus in different mouse tissues predict position of promoters (H3K4me3), enhancers (H3K27ac) and transcribed genes (H3K36me3). On the basis of H3K4me3 ChIP-seq data, only the *Myc* and *Pvt1* promoters showed activity in bone marrow cells and in the CH12 lymphoma cell line at the *Myc* locus. Consistent with the active transcription of these genes, the *Myc* and *Pvt1* gene bodies were marked by H3K36me3. Several strong H3K27ac peaks are specifically present in the BENC region in

haematopoietic tissues or derived cell lines (bone marrow, bone-marrow derived macrophages (BMDM), embryonic day (E)14.5 fetal liver (but not in adult liver)), but not in other non-blood-related samples, including brown adipose tissue (BAT), mouse embryonic stem cell line E14 (ES-E14), mouse embryonic fibroblasts (MEF), olfactory bulb (Olfact) and small intestine (Smint). Other putative enhancers, centromeric to *Myc* or overlapping with *Pvt1* are also indicated. Data are from GEO accession number GSE29184 and ref. 6.



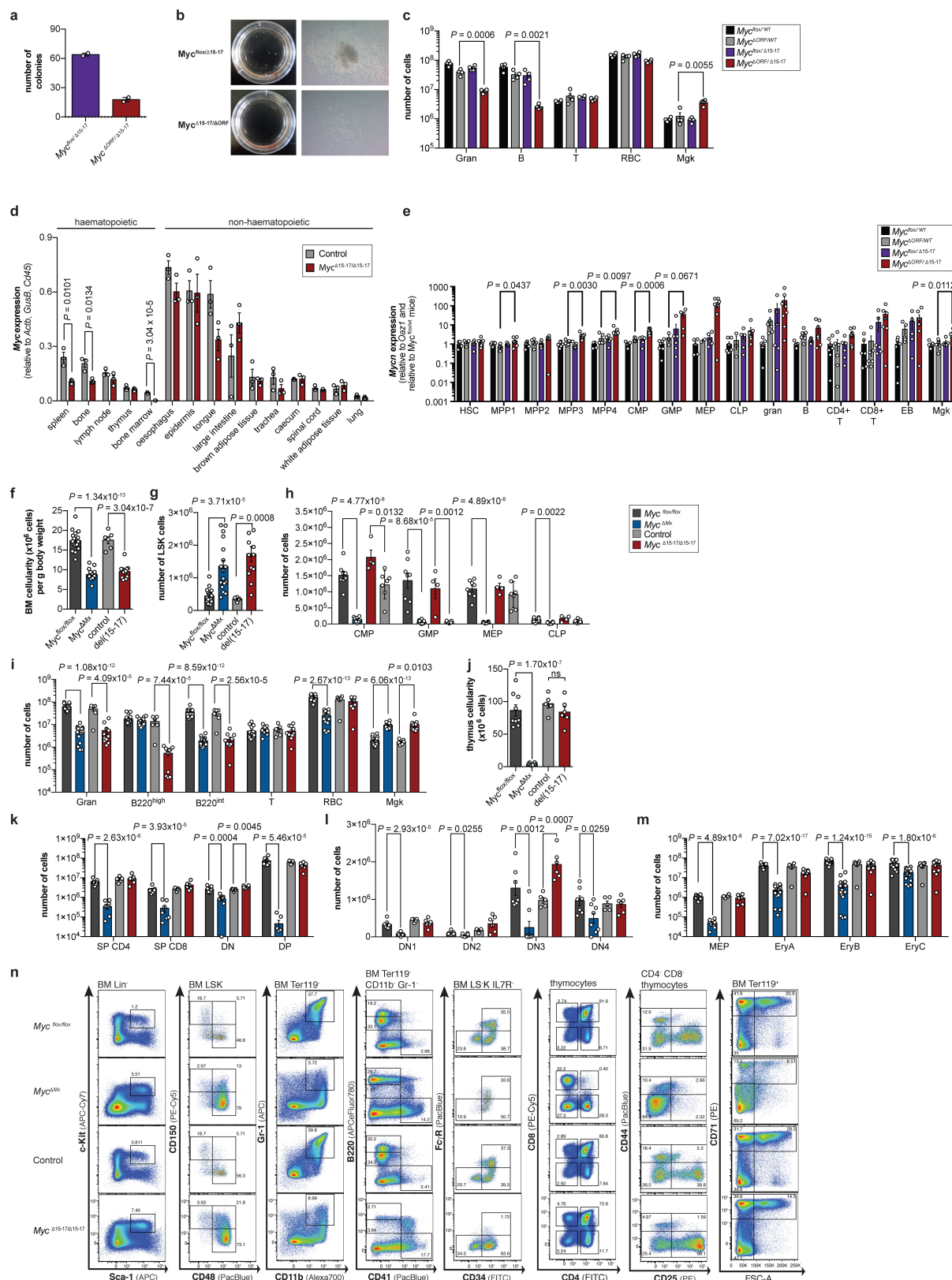
Extended Data Figure 2 | Enhancer activity of the BENC region located 1.7 Mb downstream of the *Myc* gene in haematopoietic stem and progenitor cells. **a**, Top, schematic representation of the *Myc* locus on chromosome 15 within its topologically associating domain⁴ (TAD; brown bar). The position of genes (arrows), predicted promoters (H3K4me3, blue boxes) and enhancers (H3K4me1, H3K27ac, other boxes) are shown. Bottom, representation of the different mouse transposon insertions (top) and Cre-mediated deletions (bottom, red bars) used to identify enhancer regions⁵. **b–k**, LacZ activity measured by FDG staining in bone marrow derived from LacZ-reporter mice described in **a** using flow cytometry. **b**, Representative flow cytometry histograms of the percentage of FDG⁺ cells (\pm s.d., vertical) and the geometric mean fluorescence intensity (MFI \pm s.e.m.) of HSCs, LSK cells, myeloid cells and lymphoid cells from wild-type, heterozygous 17a and Δ 15-17 mice. Data are derived from two independent experiments. **c**, LacZ activity in HSCs and MPP1-4 cells isolated from wild-type (Non-tg), heterozygous 3a, 17a

and Δ 15-17 mice and shown as geometric mean fluorescence intensity (MFI \pm s.e.m.) values. **d**, **e**, LacZ activity in HSCs and MPP1-4 cells from wild-type (Non-tg), heterozygous 3a, 17a and Δ 15-17 mice shown as geometric mean fluorescence intensity (MFI \pm s.e.m.). **f–h**, Representative histograms showing lacZ activity in HSC and MPP (**f**), progenitor (**g**) and differentiated cell (**h**) populations of bone marrow from wild-type (Non-tg) as well as heterozygous 17a and 14b mice. EB, erythroblast. Data are mean percentage of FDG⁺ cells (\pm s.d.) from two independent experiments. **i–k**, LacZ activity in HSPCs (LSK), myeloid committed progenitor (LSK⁻K) and differentiated cell (Lin⁺) populations of heterozygous mice carrying the indicated insertions or deletions measured by FDG staining. Data are geometric mean fluorescence intensity (MFI \pm s.e.m.) and mean percentage of FDG⁺ cells (\pm s.d., vertical); representative data from two independent experiments are shown. The sample size is as follows: $n = 3$ mice per group.



Extended Data Figure 3 | The enhancer region (15–17) is critical for HSC function and interacts with the *Myc* promoter. **a–d**, Comparison of control and homozygous *Myc^{Δ15-17/Δ15-17}* mice. Body weight (**a**), bone marrow cellularity normalized to body weight (**b**), number of LSK cells (**c**) and representative flow cytometry profiles using indicated markers for differentiated cell populations (**d**) are shown. **e, f**, Transplantation of homozygous *Myc^{Δ15-17/Δ15-17}* bone marrow cells in a competitive setting. Representative flow cytometry profiles showing the peripheral blood (PB) chimerism of transplanted CD45.2⁺ (either homozygous *Myc^{Δ15-17/Δ15-17}* or control) cells as indicated (**e**) and of HSC and MPP1–4 cells derived from the bone marrow of competitively transplanted mice 16 weeks after transplantation (**f**). **g, h**, Transplantation of LSK combined with LS⁺ K cells derived from homozygous *Myc^{Δ15-17/Δ15-17}* or controls into T-cell-deficient NSG mice. Distribution of CD4⁺ and CD8⁺ expressing mature T cell populations (**g**) and thymic progenitors (**h**) derived from

transplanted homozygous *Myc^{Δ15-17/Δ15-17}* or control cells. **i, j**, Physical proximity between *Myc* and *BENC* revealed by DNA FISH in HSPCs. **i**, Schematic representation of the locus, including the position of the three BACs used in DNA FISH and their relative distances. **j**, Two-dimensional projection images and three-dimensional reconstruction of double-staining DNA FISH for the BACs indicated for representative nuclei of LSK cells. **k**, Three-dimensional distance measurement of the BACs in micrometre scale. The box plot shows the 3rd quartile, median and the 1st quartile. The whiskers of the box plot extend to the data points less than $1.5 \times$ the interquartile range from the 1st and the 3rd quartile. The number of measurements for each double staining (*n*) is indicated. Sample sizes are as follows: $n = 5$ mice per group from two independent experiments (a–c); $n = 3$ mice per group from one experiment (g, h); $n = 62$ cells (397P6_22L9), $n = 55$ (22L9_207P4), $n = 34$ cells (397P6_207P4) from one experiment (k). *P* values shown are from an unpaired two-tailed *t*-test.

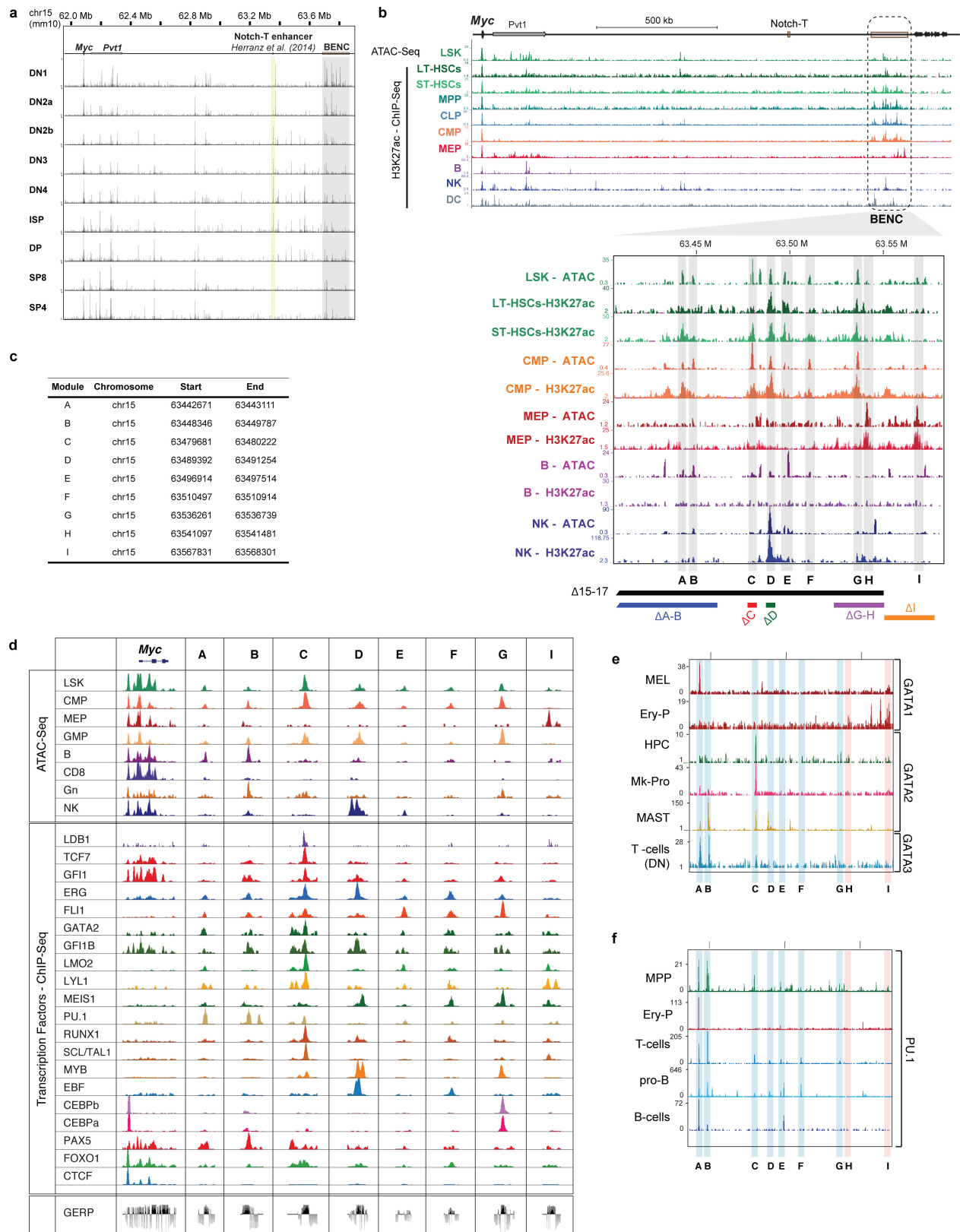


Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | The *Myc*^{Δ15-17} deletion is allelic to the *Myc*^{ΔORF} deletion and the bone marrow phenotype of *Myc*^{Δ15-17} mice closely mimics *Mx-Cre*-mediated conditional deletion of *Myc*.

a, Number of colonies obtained in colony-forming assays using bone marrow cells from the indicated genotypes. Mean values (\pm s.e.m.) of two biological replicates (dots) with technical duplicates are shown. **b**, Pictures of representative dishes (left) and colonies (right) of one experiment with two biological duplicates and technical duplicates. **c**, Total cell numbers of differentiated cell types present in different compound heterozygous mice isolated from legs, hips and spine. **d**, Relative *Myc* expression in haematopoietic and non-haematopoietic tissues obtained from homozygous *Myc*^{Δ15-17/Δ15-17} and control mice. Data are mean \pm s.e.m. of three mice. **e**, Relative *Mycn* mRNA expression in haematopoietic bone marrow cell populations derived from indicated mutant mice. All data are mean \pm s.e.m. **f–n**, Comparison of adult mice with poly(I:C)-induced *Mx-Cre*-mediated deletion of the *Myc* gene with ones carrying a homozygous *Myc*^{Δ15-17/Δ15-17} allele. Bone marrow cellularity normalized to body weight (**f**), number of LSK cells (**g**), number of committed progenitor populations (**h**), number of differentiated cells (**i**), thymus cellularity (**j**), number of thymic mature T cells and progenitors (**k**) and thymic double negative (DN) populations (**l**), MEPs and erythroid progenitors (**m**) is shown. Bone marrow cell numbers refer to cells isolated from legs, hips and spine. **n**, Representative flow cytometry profiles of bone marrow cells derived from mouse mutants indicated on the left and gated as indicated at the top stained with indicated cell surface markers. All data are mean \pm s.e.m. Sample sizes are as follows: $n = 2$ mice per group analysed each in technical duplicates (**a**); $n = 4$ mice per group

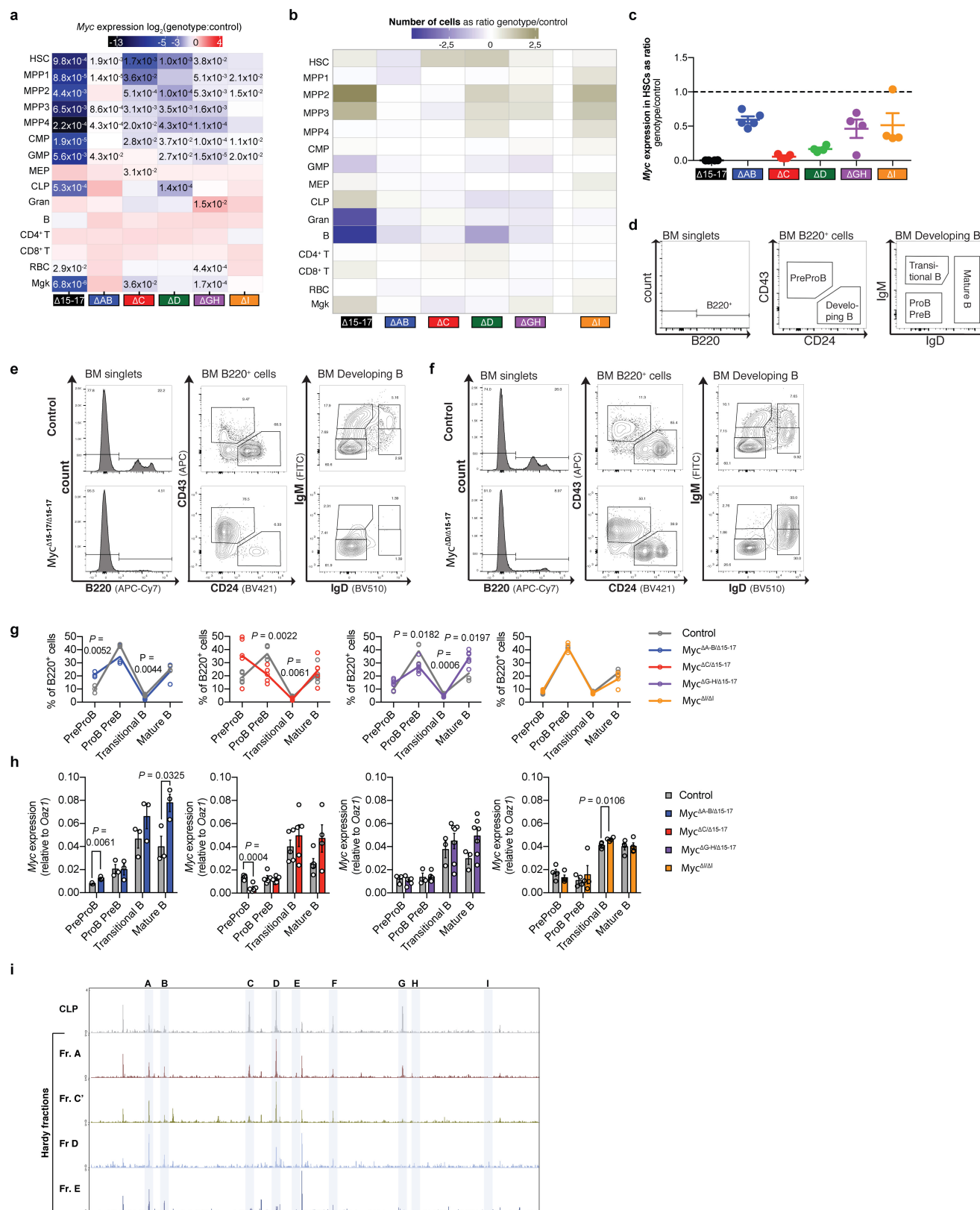
from one experiment (**c**); $n = 3$ mice per group from one experiment (**d**); *Myc*^{WT/flox} ($n = 7$ mice for all populations, except MPP1 ($n = 6$ mice), MEP ($n = 4$ mice), CD8⁺ T ($n = 6$ mice), megakaryocytes (Mgk) ($n = 6$ mice)); *Myc*^{ΔORF/WT} ($n = 6$ mice for all populations, except HSCs ($n = 5$ mice), CMP ($n = 5$ mice), GMP ($n = 3$ mice), MEP ($n = 2$ mice), CLP ($n = 3$ mice), RBC ($n = 5$ mice)); *Myc*^{Δ15-17/flox} ($n = 6$ mice for all populations, except MPP1 ($n = 5$ mice), MPP3 ($n = 5$ mice), MEP ($n = 5$ mice), RBC ($n = 5$ mice), megakaryocytes ($n = 5$)); *Myc*^{Δ15-17/ΔORF} ($n = 8$ mice for all populations, except MPP1 ($n = 7$ mice), MPP3 ($n = 7$ mice), GMP ($n = 5$ mice), CLP ($n = 7$ mice), CD8⁺ T ($n = 7$ mice)), data from two independent experiments (**e**); $n = 17$ mice for *Myc*^{flox/flox} and *Myc*^{ΔMx}, $n = 6$ mice for control, $n = 12$ mice for *Myc*^{Δ15-17/Δ15-17}, from two independent experiments (**f**, **g**); $n = 8$ mice for *Myc*^{flox/flox} and *Myc*^{ΔMx}, $n = 4$ mice for control, $n = 7$ mice for *Myc*^{Δ15-17/Δ15-17}, from two independent experiments (**h**); $n = 17$ mice for *Myc*^{flox/flox} and *Myc*^{ΔMx}, $n = 6$ mice for control and $n = 11$ mice for *Myc*^{Δ15-17/Δ15-17}, from two independent experiments (**i**); $n = 9$ mice for *Myc*^{flox/flox}, $n = 8$ mice for *Myc*^{ΔMx}, $n = 5$ mice for control and $n = 6$ mice for *Myc*^{Δ15-17/Δ15-17} from one experiment (**j**); $n = 7$ mice for *Myc*^{flox/flox}, $n = 8$ mice for *Myc*^{ΔMx}, $n = 5$ mice for control and $n = 6$ mice for *Myc*^{Δ15-17/Δ15-17}, from two independent experiments (**k**, **l**); $n = 7$ (MEP) and $n = 17$ (all other populations) mice for *Myc*^{flox/flox} and *Myc*^{ΔMx}, $n = 4$ (MEP) and $n = 6$ mice (all other populations) for control, $n = 7$ (MEP) and $n = 11$ (all other populations) mice for *Myc*^{Δ15-17/Δ15-17}, from two independent experiments (**m**); two independent experiments, except for the thymus (one experiment) (**n**). *P* values are from unpaired two-tailed *t*-test.



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | BENC is a multi-modular enhancer and recruits haematopoietic transcription factors to its constituents in a cell-type-specific manner. **a**, Overview of ATAC-seq profiles from various thymic progenitor populations in the *Myc* locus (data from ImmGen repository⁴⁸). The Notch-responsive enhancer described in ref. 9 is highlighted in yellow and BENC in grey. **b**, Top, overview of the *Myc* locus and adjacent regions including BENC. The ATAC-seq profile of LSK cells and H3K27ac profiles for various haematopoietic cell populations are shown. BENC is clearly marked in a cell-type-specific manner by H3K27ac and the chromatin accessible in LSK cells as measured by ATAC-seq. Bottom, same graphic as shown in Fig. 3a. ATAC-seq and H3K27ac profiles of various haematopoietic cell types. Both the accessibility and deposition of H3K27ac change in a cell-type-specific manner. **c**, Genomic coordinates of BENC modules in the mouse genome (genome assembly mm9). **d**, Top, ATAC-seq profiles reveal an open chromatin configuration at the *Myc* promoter in all blood cells tested,

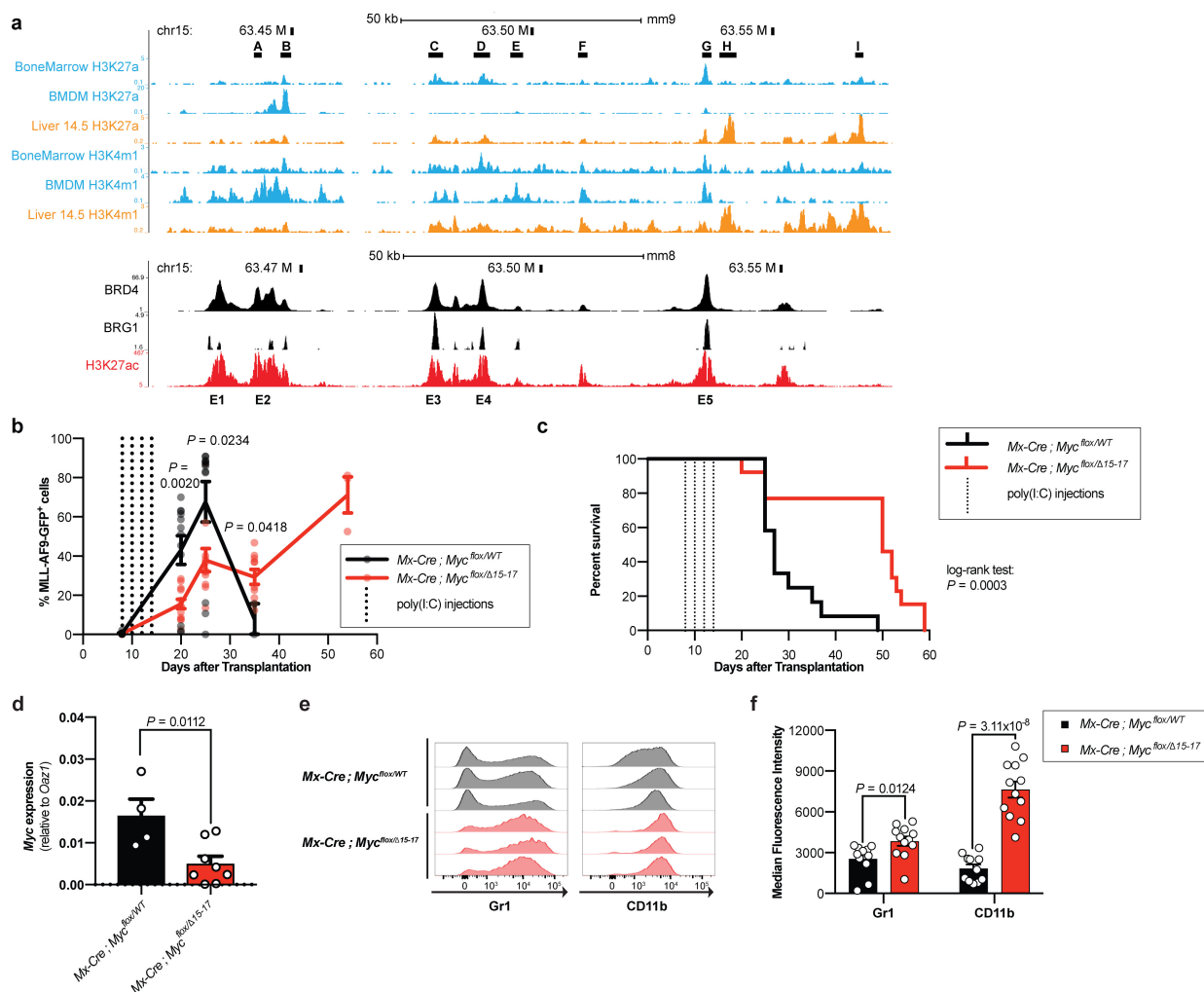
whereas chromatin accessibility at the different BENC modules is much more dynamic. For example, ATAC peaks in LSK cells were present at the C, D and G modules, whereas MEPs showed a double peak at module I and natural killer (NK) cells showed a double peak at module D. Bottom, chromatin immunoprecipitation followed by sequencing (ChIP-seq) profiles of transcription factors aligned to the *Myc* promoter and BENC modules A–I. Several important haematopoietic transcriptional regulators are not detected at the *Myc* promoter or only bind faintly to it (MEIS1, FLI1, PU.1, RUNX1, SCL (also known as TAL1), MYB), whereas they bind strongly to BENC modules in a differential manner. The ChIP-seq data were extracted from different sources outlined in Extended Data Table 1. **e**, ChIP-seq profiles for GATA1, -2, -3 in different haematopoietic cells (see Extended Data Table 1 for sources). Ery-P, erythrocyte progenitors; MPP, multipotent progenitor; HPC, haematopoietic progenitor cell. **f**, ChIP-seq profiles of PU.1 showing preferential occupancy at the A and B modules but weak signals at the other BENC modules.



Extended Data Figure 6 | See next page for caption.

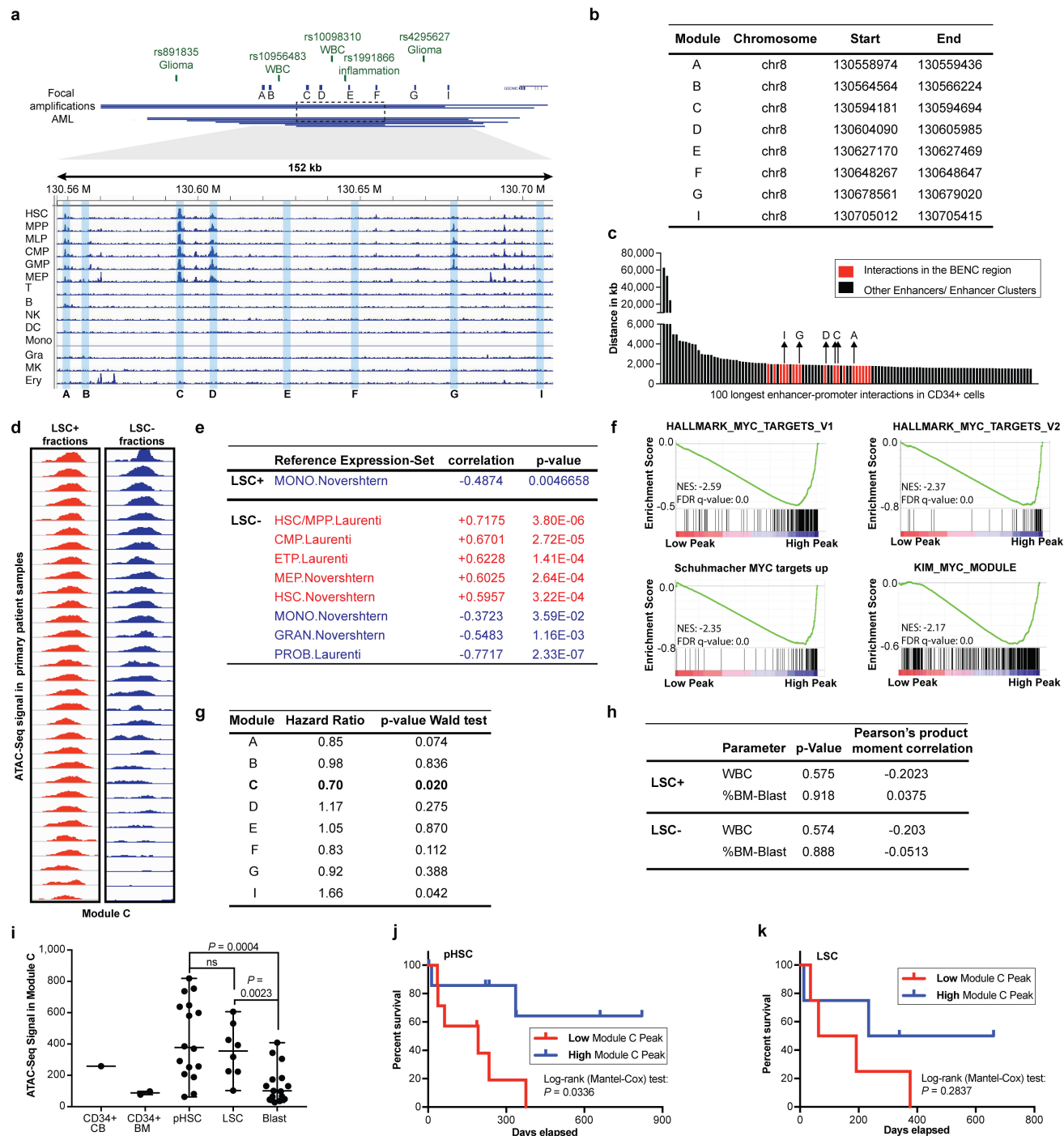
Extended Data Figure 6 | Consequences of deletion of individual BENC modules for bone marrow populations and during B cell development. **a**, Copy of Fig. 3c with exact *P* values from unpaired two-tailed *t*-tests included in the heat map tiles. **b**, Changes in cell numbers of various bone marrow haematopoietic cell populations in *Myc^{Δmodule/Δ15-17}* mice. Data are shown as log₂-transformed mean values of the ratio of *Myc^{Δmodule/Δ15-17}* mice to the respective controls. **c**, *Myc* expression in HSCs in homozygous *Myc^{Δ15-17/Δ15-17}* and *Myc^{Δmodule/Δ15-17}* mice. **d**, Flow cytometric gating strategy used to quantify PreProB (B220⁺CD24⁻CD43⁺), ProB PreB (B220⁺CD24⁺CD43⁺IgM⁻IgD⁻), transitional B (B220⁺CD24⁺CD43⁻IgM⁺IgD⁻), and mature B (B220⁺CD24⁺CD43⁻IgD⁺) cells. **e**, **f**, Representative expression profiles of homozygous *Myc^{Δ15-17/Δ15-17}* (**e**) and *Myc^{ΔD/Δ15-17}* (**f**) mice analysed by flow cytometry. Data show a reduction in the fraction of B220⁺ cells as well as an accumulation of PreProB cells in both mutants. **g**, **h**, Cell frequencies and *Myc* expression during early B cell development in *Myc^{Δmodule/Δ15-17}* mice showing effects of module A–B, C, G–H and I deletion. **g**, Quantification of PrePro B, ProB PreB, transitional B and mature B cells in mice with the indicated BENC module deletions (*Myc^{Δmodule/Δ15-17}*) shown as the frequencies of B220⁺ cells. **h**, Relative *Myc* mRNA levels in early B cell developmental stages in mice with BENC module deletions. Data are mean ± s.e.m. **i**, ATAC-seq profiles of BENC in CLPs and B cell progenitors obtained from the Immgen⁴⁸ repository. Sample sizes are as follows. **a**, See sample sizes for Fig. 3c in the Methods. **b**, *Myc^{Δ15-17/WT}* control for *Myc^{Δ15-17/Δ15-17}* mice (*n* = 4 mice for all other populations, except HSC, MPP1, MPP3, MEP (*n* = 3 mice), *Myc^{Δ15-17/Δ15-17}* mice (*n* = 3 mice for all populations), *Myc^{Δ15-17/WT}* control for *Myc^{Δ15-17/ΔA-B}* mice (*n* = 10 mice), *Myc^{Δ15-17/ΔA-B}* mice (*n* = 10 mice), *Myc^{Δ15-17/WT}* control for *Myc^{Δ15-17/ΔC}* mice (*n* = 14 for all

populations, except HSCs, MPP1, MPP2, MPP3 and MPP4 (*n* = 12 mice)), *Myc^{Δ15-17/ΔC}* mice (*n* = 14 mice for all populations, except B, granulocytes, RBC, megakaryocytes, CD4⁺ T and CD8⁺ T cells (*n* = 13 mice)), *Myc^{Δ15-17/WT}* control for *Myc^{Δ15-17/ΔD}* mice (*n* = 14 mice for all populations, except HSCs, MPP1, MPP2, MPP3, MPP4 (*n* = 12 mice)), *Myc^{Δ15-17/ΔD}* mice (*n* = 9 mice for all populations, except HSCs, MPP1, MPP2, MPP3, MPP4 (*n* = 6 mice), *Myc^{Δ15-17/WT}* control for *Myc^{Δ15-17/ΔG-H}* mice (*n* = 8 mice for all populations, except HSCs, MPP1, MPP2, MPP3, MPP4, CMP, GMP, MEP, CLP (*n* = 6 mice)), *Myc^{Δ15-17/ΔG-H}* mice (*n* = 12 mice for all populations, except HSCs, MPP1, MPP2, MPP3, MPP4, CMP, GMP, MEP, CLP (*n* = 7 mice)), *Myc^{WT/WT}* control for *Myc^{ΔI/ΔI}* mice (*n* = 9 mice for all populations), *Myc^{ΔI/ΔI}* mice (*n* = 9 mice) from two to three independent experiments. **c**, See sample sizes for Fig. 3c in the Methods. **e**, **f**, See sample sizes for Fig. 3e in the Methods. **g**, *Myc^{Δ15-17/WT}* control for *Myc^{Δ15-17/ΔA-B}* mice (*n* = 3 mice), *Myc^{Δ15-17/ΔA-B}* mice (*n* = 3 mice), *Myc^{Δ15-17/WT}* control for *Myc^{Δ15-17/ΔC}* mice (*n* = 5 mice), *Myc^{Δ15-17/ΔC}* mice (*n* = 5 mice), *Myc^{Δ15-17/WT}* control for *Myc^{Δ15-17/ΔG-H}* mice (*n* = 3 mice), *Myc^{Δ15-17/ΔG-H}* mice (*n* = 7 mice), *Myc^{WT/WT}* control for *Myc^{ΔI/ΔI}* mice (*n* = 4 mice), *Myc^{ΔI/ΔI}* mice (*n* = 4 mice), data from one experiment. **h**, *Myc^{Δ15-17/WT}* control for *Myc^{Δ15-17/ΔA-B}* mice (*n* = 3 mice), *Myc^{Δ15-17/ΔA-B}* mice (*n* = 3 mice), *Myc^{Δ15-17/WT}* control for *Myc^{Δ15-17/ΔC}* mice (*n* = 5 mice), *Myc^{Δ15-17/ΔC}* mice (*n* = 5 mice for all populations, except mature B cells (*n* = 4 mice)), *Myc^{Δ15-17/WT}* control for *Myc^{Δ15-17/ΔG-H}* mice (*n* = 3 mice), *Myc^{Δ15-17/ΔG-H}* mice (*n* = 7 mice for all populations, except PrePro B and ProB PreB (*n* = 6 mice)), *Myc^{WT/WT}* control for *Myc^{ΔI/ΔI}* mice (*n* = 4 mice), *Myc^{ΔI/ΔI}* mice (*n* = 4 mice), data from one experiment. *P* values are from an unpaired two-tailed *t*-test.



Extended Data Figure 7 | The BENC enhancer modules overlap with a mouse leukaemia super-enhancer and loss of BENC in an MLL-AF9 leukaemia mouse model prolongs survival. **a**, Comparison of BENC modules from normal haematopoietic tissues (see also Extended Data Fig. 1) with super-enhancer elements as defined previously². Localization of BENC modules is shown above ChIP-seq tracks and super-enhancer elements defined by Brd4 occupancy as well as broad distribution of H3K27ac marks are indicated at the bottom. **b, c**, Loss of BENC in MLL-AF9-mediated leukaemias (for experimental setup, see Fig. 4a) in response to poly(I:C) injections (dotted lines) leads to a delay in AML progression (**b**) and to an increased survival of the mice that initially had leukaemia (**c**). In order to induce Cre expression from the *Mx1* promoter, mice transplanted with leukaemic cells were subjected to four injections of poly(I:C) starting eight days after transplantation. In contrast to the experiment shown in Fig. 4b, c, mice were thereafter not injected with additional rounds of poly(I:C). As a consequence, leukaemic cells that had not recombined (genomic escapees) survived and expanded as Myc-expressing *Mx-Cre;Myc^{Δ15-17/flox}* cells, causing the death of the recipient mice. Together with the data presented in Fig. 4b, c, in which continuous injections of poly(I:C) result in the clearance of leukaemic cells from

the peripheral blood in some mice, this argues for an insufficient deletion of the conditional *Myc* allele in this experiment and demonstrates that BENC is essential for maintenance of leukaemia. **d**, Loss of BENC in MLL-AF9 leukaemic cells isolated from the peripheral blood results in a significant reduction in *Myc* expression. **e, f**, Upregulation of myeloid differentiation markers on blast cells in the peripheral blood after *Mx-Cre*-mediated deletion of BENC. **e**, Three representative histogram plots of Gr1 and CD11b expression for leukaemic blasts from peripheral blood of *n* = 11 mice (*Mx-Cre;Myc^{WT/flox}*) or *n* = 12 mice (*Mx-Cre;Myc^{Δ15-17/flox}*). **f**, Quantification of Gr1 and CD11b expression of blasts. MFI is shown as mean ± s.e.m. Sample sizes are as follows: *Mx-Cre;Myc^{WT/flox}* (*n* = 12 mice for 8, 20 and 25 days, *n* = 2 mice for 35 days), *Mx-Cre;Myc^{Δ15-17/flox}* (*n* = 13 mice for 8 days, *n* = 12 mice for 20 days, *n* = 10 mice for 25 days, *n* = 9 mice for 35 days, *n* = 3 mice for 54 days) (**b**); *n* = 12 mice for *Mx-Cre;Myc^{WT/flox}* and *n* = 13 mice for *Mx-Cre;Myc^{Δ15-17/flox}* (**c**); *n* = 4 mice for *Mx-Cre;Myc^{WT/flox}* and *n* = 8 mice for *Mx-Cre;Myc^{Δ15-17/flox}* (**d**); *n* = 11 mice for *Mx-Cre;Myc^{WT/flox}* and *n* = 12 mice for *Mx-Cre;Myc^{Δ15-17/flox}* (**f**). *P* values are from an unpaired two-tailed *t*-test (**b, d, f**) or two-tailed Wilcoxon rank-sum test (**c**).



Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | The BENC structure is conserved in the human genome and module C is differentially regulated in LSCs and its accessibility correlates with overall patient survival. **a**, ATAC-seq analysis of human haematopoietic cell types in the BENC region. Risk single-nucleotide polymorphisms that are associated with haematological traits and focal amplifications in patients with AML are shown at the top^{22–25}. **b**, Genomic coordinates of BENC modules in the human genome (hg19). **c**, Display of the 100 longest *in cis* interactions of promoters with enhancers and enhancer clusters in CD34⁺ cells measured by promoter capture high-resolution chromosome conformation capture (Hi-C)¹⁴. Highlighted in red are the interactions mapped to the BENC region and interactions within BENC modules are labelled accordingly. **d**, ATAC-seq profiles of module C in primary human AML samples divided into LSC⁺ (red) and LSC[−] (blue) fractions, ranked from high to low. **e**, PERT model estimates of relative proportions of cell-type-specific transcriptional programs (MONO, monocytes; ETP, early T cell progenitor; GRAN, granulocytes; PROB, pro-B cells) composing the global gene expression of fractions. Spearman's rank correlation between these estimated cell-specific proportions of transcriptional programs and module C peaks of fractions were determined. **f**, Gene set enrichment analysis of MYC target signatures in LSC[−] fractions stratified according to ATAC-seq peak height

in module C. **g**, Correlation between the maximum peak height in module C (bold) and the other modules in LSCs and patient survival. Hazard ratios and *P* values from the Wald test are shown. Module C is the only module that is more accessible in LSC⁺ cells compared to LSC[−] cells (see Fig. 4d) and that shows a correlation with patient survival. **h**, Correlation between the ATAC-seq peak height in module C and either white blood cell counts (WBC) or percentage of bone marrow blast counts (%BM-Blast). **i–k**, Correlation of ATAC-seq signals in immunophenotypic pre-leukaemic HSCs (pHSCs), LSCs and blasts with overall survival using the GSE74912 dataset²⁹. **i**, ATAC-seq signal in module C in CD34⁺ cord blood cells (CD34⁺ CB), CD34⁺ bone marrow cells (CD34⁺ BM), pHSCs, LSCs and blasts. **j, k**, Kaplan–Meier representation of overall survival according to ATAC-seq signal in module C in pre-leukaemic HSC (**j**) and LSC (**k**) fractions. For this stratification, the patient cohort was split according to the median of the maximum ATAC-seq peak height in module C. Sample sizes are as follows: *n* = 1 patient (CD34⁺ cord blood cells), *n* = 2 bone marrow samples (CD34⁺ bone marrow cells), *n* = 16 bone marrow samples (pre-leukaemic HSCs), *n* = 8 bone marrow samples (LSCs), *n* = 15 bone marrow samples (blast) (**i**); *n* = 15 patients (**j**); *n* = 8 patients (**k**). *P* values in (**i**) from unpaired two-tailed *t*-test.

Extended Data Table 1 | Sources of publically available ChIP-seq, ATAC-seq and capture Hi-C datasets

Transcription Factor	Cell-type	Source (GEO/ ArrayExpress)
TCF7 *	Haematopoietic progenitors (Lin-Sca1+CD34+)	GSM773994
LMO2 *	Multipotent myeloid progenitor (HPC-7)	GSM552237
LDB1 *	Haematopoietic progenitor cells (Lin-)	GSM641909
GATA2 *,†	Multipotent myeloid progenitor (HPC-7)	GSM552234
SCL/TAL1 *	Haematopoietic progenitor cells	GSM641910
LYL1 *	Multipotent myeloid progenitor (HPC-7)	GSM552238
RUNX1 *	Multipotent myeloid progenitor (HPC-7)	GSM552241
ERG *	Multipotent myeloid progenitor (HPC-7)	Wilson et al. (2010)
GFI1B *	Multipotent myeloid progenitor (HPC-7)	GSM552235
EBF1 *	Pro B cells	GSM499030
MYB *	Myeloid progenitors	GSE22095
MEIS1 *	Multipotent myeloid progenitor (HPC-7)	GSM552239
FLI1 *	Multipotent myeloid progenitor (HPC-7)	GSM552233
CEBPB *	Macrophages	GSM537985
CEBPA *	Macrophages	GSM537984
SPI1/PU.1 *	Multipotent myeloid progenitor (HPC-7)	GSM552240
GATA3 *,#	T cells	GSM523221
GATA1 *,#	MEL (leukaemia cell line)	GSM912907
FOXO1 *	Pro-B cells	GSM546525
CTCF *	Multipotent myeloid progenitor (HPC-7)	GSM1167572
PAX5 *	Pre-B cells	GSM860927
SPI1/PU.1 †	Erythroid progenitors	GSE21953
SPI1/PU.1 †	T cells	GSM774291
SPI1/PU.1 †	Pro-B cells	GSM539537
SPI1/PU.1 †	B cells	GSM537989
GATA1 †	ES-cell derived erythroid progenitors	GSM867156
GATA2 †	Bone marrow haematopoietic progenitor cells	GSM641911
GATA2 †	Megakaryocyte progenitor cell line	GSM777091
GATA2 †	MAST	HAEMCODE 3
BRD4 ‡	RN2 cells	GSM1262345
BRG1 ‡	RN2 cells	GSM1262346
H3K27ac ‡	RN2 cells	GSM1262348
Capture Hi-C §	Human CD34+ stem cells	E-MTAB-2323
ATAC-Seq	Fractionated AML patient samples	GSE74912
ATAC-Seq ¶	Thymic progenitors	GSE100738
ATAC-Seq #	B cell progenitors	GSE100738

*Used for Fig. 4b and Extended Data Fig 5d.

†Used for Extended Data Fig. 5e.

‡Used for Extended Data Fig. 7a.

§Used for Extended Data Fig. 8c.

||Used for Extended Data Fig. 8i.

¶Used for Extended Data Fig. 5a.

#Used for Extended Data Fig. 6i.

Extended Data Table 2 | Cell-surface phenotypes of analysed cell populations

Abbreviation	Population name	Cell surface phenotype
Lin	lineage	CD11b Gr-1 B220 CD4 CD8a Ter119
HSPC/ LSK	haematopoietic stem and progenitor cells	Lin- Sca-1+ c-Kit+
LS-K	Haematopoietic progenitor cells	Lin- Sca-1- c-Kit+
HSC	haematopoietic stem cell	LSK CD150+ CD48- CD34- CD135-
MPP1	multipotent progenitor 1	LSK CD150+ CD48- CD34+ CD135-
MPP2	multipotent progenitor 2	LSK CD150+ CD48+ CD34+ CD135-
MPP3	multipotent progenitor 3	LSK CD150- CD48+ CD34+ CD135-
MPP4	multipotent progenitor 4	LSK CD150- CD48+ CD34+ CD135+
CMP	common myeloid progenitor	Lin- Sca-1- c-Kit+ IL7Ra- CD34+ FcgRlow
GMP	granulocyte macrophage progenitor	Lin- Sca-1- c-Kit+ IL7Ra- CD34+ FcgRhigh
MEP	megakaryocyte erythrocyte progenitor	Lin- Sca-1- c-Kit+ IL7Ra- CD34- FcgRlow
CLP	common lymphoid progenitor	Lin- Sca-1int c-Kitint IL7Ra+
Gran	granulocytes	CD71- Ter119- CD11b- Gr-1+
B	B lymphocytes	CD71- Ter119- CD11b- Gr-1- B220+
B220high	B220high lymphocytes	CD71- Ter119- CD11b- Gr-1- B220high
B220int	B220int lymphocytes	CD71- Ter119- CD11b- Gr-1- B220intermediate
PreProB	PreProB lymphocytes	B220+ CD24- CD43+
ProB PreB	ProB PreB lymphocytes	B220+ CD24+ CD43- IgM- IgD-
Trans B	Transitional B lymphocytes	B220+ CD24+ CD43- IgM+ IgD-
mature B	mature B lymphocytes	B220+ CD24+ CD43- IgD+
T	T lymphocytes	CD71- Ter119- CD11b- Gr-1-; CD4+ cells and CD8a+ cells
CD4+ T	CD4+ T lymphocytes	CD71- Ter119- CD11b- Gr-1- CD4+
CD8+ T	CD8+ T lymphocytes	CD71- Ter119- CD11b- Gr-1- CD8a+
DP	double positive thymocytes	CD4+ CD8a+
SP CD4	single positive CD4 thymocytes	CD4+
SP CD8	single positive CD8 thymocytes	CD8a+
DN	double negative thymocytes	CD4- CD8a-
DN1	DN1 thymocytes	CD4- CD8a- CD44+ CD25-
DN2	DN2 thymocytes	CD4- CD8a- CD44+ CD25+
DN3	DN3 thymocytes	CD4- CD8a- CD44- CD25+
DN4	DN4 thymocytes	CD4- CD8a- CD44- CD25-
EB	erythroblasts	CD71+ Ter119-
EP	erythrocyte progenitors	CD71+
EryA	erythrocyte progenitor A	CD71+ Ter119+ FSChigh
EryB	erythrocyte progenitor B	CD71+ Ter119+ FSClow
EryC	erythrocyte progenitor C	CD71- Ter119+ FSClow
RBC	red blood cells	Ter119+
Mgk	megakaryocytes	CD71- Ter119- CD11b- Gr-1- CD41+

Structure and mutagenesis reveal essential capsid protein interactions for KSHV replication

Xinghong Dai^{1,2,3*}, Danyang Gong^{3*}, Hanyoung Lim¹, Jonathan Jih¹, Ting-Ting Wu³, Ren Sun^{2,3} & Z. Hong Zhou^{1,2}

Kaposi's sarcoma-associated herpesvirus (KSHV) causes Kaposi's sarcoma^{1,2}, a cancer that commonly affects patients with AIDS³ and which is endemic in sub-Saharan Africa⁴. The KSHV capsid is highly pressurized by its double-stranded DNA genome, as are the capsids of the eight other human herpesviruses⁵. Capsid assembly and genome packaging of herpesviruses are prone to interruption^{6–9} and can therefore be targeted for the structure-guided development of antiviral agents. However, herpesvirus capsids—comprising nearly 3,000 proteins and over 1,300 Å in diameter—present a formidable challenge to atomic structure determination¹⁰ and functional mapping of molecular interactions. Here we report a 4.2 Å resolution structure of the KSHV capsid, determined by electron-counting cryo-electron microscopy, and its atomic model, which contains 46 unique conformers of the major capsid protein (MCP), the smallest capsid protein (SCP) and the triplex proteins Tri1 and Tri2. Our structure and mutagenesis results reveal a groove in the upper domain of the MCP that contains hydrophobic residues that interact with the SCP, which in turn crosslinks with neighbouring MCPs in the same hexon to stabilize the capsid. Multiple levels of MCP–MCP interaction—including six sets of stacked hairpins lining the hexon channel, disulfide bonds across channel and buttress domains in neighbouring MCPs, and an interaction network forged by the N-lasso domain and secured by the dimerization domain—define a robust capsid that is resistant to the pressure exerted by the enclosed genome. The triplexes, each composed of two Tri2 molecules and a Tri1 molecule, anchor to the capsid floor via a Tri1 N-anchor to plug holes in the MCP network and rivet the capsid floor. These essential roles of the MCP N-lasso and Tri1 N-anchor are verified by serial-truncation mutageneses. Our proof-of-concept demonstration of the use of polypeptides that mimic the smallest capsid protein to inhibit KSHV lytic replication highlights the potential for exploiting the interaction hotspots revealed in our atomic structure to develop antiviral agents.

We purified intact KSHV virions and used electron-counting cryo-electron microscopy (cryo-EM) to obtain a 4.2 Å resolution capsid structure (Fig. 1a, Extended Data Figs 1–4, Supplementary Video 1). The *T* = 16 icosahedral capsid contains MCP pentamers (pentons) and hexamers (hexons) decorated with the smallest capsid protein (SCP), and joined by heterotrimeric triplexes (Ta–Tf) composed of a Tri1 and two Tri2 proteins (Fig. 1b). We built atomic models for a total of 46 unique conformers of the four capsid proteins: 15 hexon MCPs, 1 penton MCP, 15 hexon SCPs, 5 Tri1 proteins and 10 Tri2 proteins (Fig. 1c, d, Extended Data Table 1, Extended Data Fig. 2d), with approximately 26,000 amino acid residues in total.

The 1,376-amino-acid MCP subunit is L-shaped: it is hinged, with a 'tower' and a 'floor' component (Fig. 2a, Supplementary Video 2). The tower contains the upper, channel and buttress domains, and the floor is made up of the bacteriophage HK97-like domain (the Johnson fold)¹¹, helix-hairpin domain, dimerization domain and N-lasso domain.

The Johnson fold is characterized by a central five-stranded β -sheet, elaborated by a long 'spine helix' and an extended 'E-loop' (Fig. 2j, k). It was first discovered in bacteriophage HK97 capsid protein gp5 (ref. 12), and subsequently in many other bacteriophages^{13–16}, the herpes simplex virus type 1 (HSV-1)¹⁷, and even bacteria and archaea^{18,19}, though the topology of the joining of the strands of the central β -sheet varies between these groups¹⁴. The KSHV Johnson-fold domain possesses the same strand-joining topology as HK97 gp5 (Fig. 2j, k).

Among the capsid proteins of the more than 100 known herpesviruses, the only published atomic structure is that of the HSV-1 MCP upper domain (MCPud) determined by X-ray crystallography²⁰. The structures of MCPud from KSHV and HSV-1 are similar²¹, as are the interactions between neighbouring MCPuds mediated by the 'major helix' (yellow in Fig. 2b). However, a loop (amino acids 767–781) in the HSV-1 MCPud is replaced by a helix (amino acids 763–778, cyan in Extended Data Fig. 5a) in KSHV MCP, which produces a groove into which an SCP binds. Our atomic model of SCP consists of an N-terminal loop, two short helices, a stem helix and a bridging helix, and is folded like the treble clef symbol (Fig. 1d). The stem helix binds the groove of one MCP and, in hexons, the bridging helix crosslinks to a neighbouring MCP (Fig. 2b, c). Our previous SCP-deletion mutagenesis of KSHV demonstrated the role of SCP crosslinking in the production of virions that contain DNA⁸. Although there is no SCP bound to penton MCP in HSV-1 (ref. 22), we clearly observed an SCP bound to each penton MCP in KSHV, albeit without crosslinking. Our model indicates that MCP–SCP interactions involve a series of hydrophobic residues (such as Tyr774, Val839, Phe840 and Leu883) on MCP. In particular, Phe840 of MCP inserts into a hydrophobic pocket formed by Leu15, Leu24, Val25, Leu49, Leu52 and Ile53 of SCP; Tyr774 of MCP has an aromatic–aromatic interaction (π -stacking) with Phe51 of SCP (Extended Data Fig. 5a). Using co-immunoprecipitation, we investigated point mutations expressed in the KSHV MCPud (amino acids 478–1,033) that replaced these hydrophobic residues with hydrophilic residues; this abolished *in vitro* MCP–SCP interactions. However, in three other residues that are not in close proximity to SCP, the same type of point mutations did not affect MCP–SCP binding (Extended Data Fig. 5b). This corroboration of our structural and functional data establishes the key residues involved in MCP–SCP binding that can be targeted for intervention (Fig. 2d–f).

The MCP buttress domain provides a vertical architectural support for the MCPud and a horizontal architectural support for the MCP channel domain that constricts the capsomer channel (Fig. 2a). The buttress domain features a four-stranded β -sheet and a helix-rich periphery (Fig. 2g, h). The channel domain fashions a six-stranded β -sheet, which is flanked by three channel-lining hairpins and by a three-helix arm that latches onto the buttress domain (Fig. 2h, i). Within each MCP, the β -sheets from the buttress and the channel domains form a β -sandwich (Fig. 2h). Across adjacent MCPs in a hexon, the six-stranded β -sheet of one channel domain is augmented by a β -strand from a channel domain

¹Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles (UCLA), Los Angeles, California 90095, USA. ²The California NanoSystems Institute (CNSI), University of California, Los Angeles (UCLA), Los Angeles, California 90095, USA. ³Department of Molecular and Medical Pharmacology, University of California, Los Angeles (UCLA), Los Angeles, California 90095, USA.

*These authors contributed equally to this work.

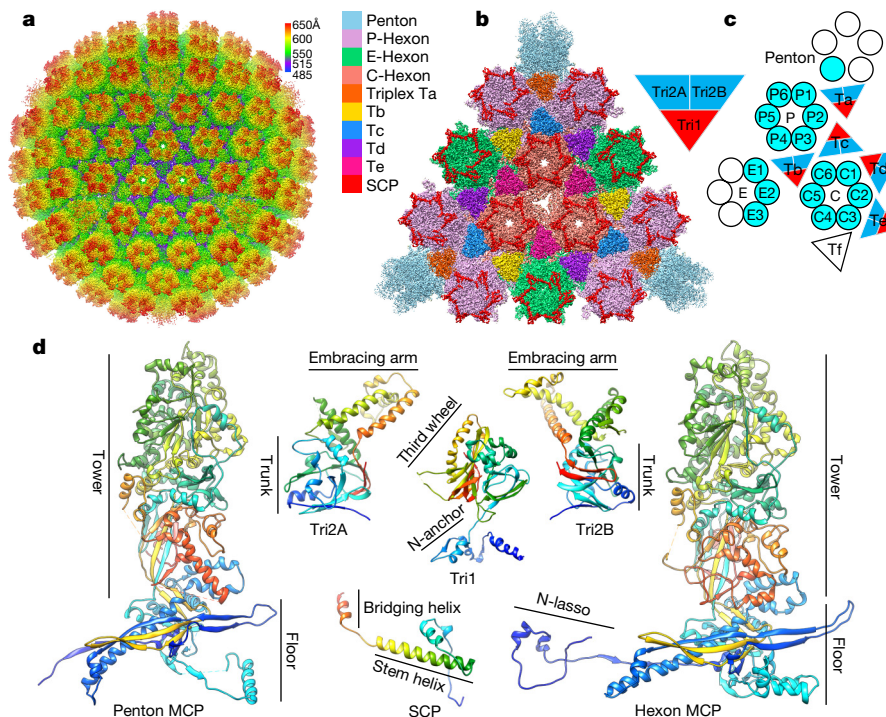


Figure 1 | Cryo-EM reconstruction and atomic modelling of KSHV capsid. **a**, Radially-coloured cryo-EM density map of KSHV capsid viewed along a three-fold axis. **b**, Zoomed in view of one facet of the icosahedral capsid. Densities of triplex Tf and penton SCP are not displayed.

c, Schematic of one asymmetric unit (shaded) of the capsid. The inset illustrates the heterotrimeric nature of a triplex. **d**, Atomic models of individual capsid proteins in ribbon representation in rainbow colours, from the N terminus (blue) to the C terminus (red).

of an adjacent MCP; this β -strand is also connected via a loop to the six-stranded β -sheet of this adjacent MCP. In this way, six β -sheets are joined together to form a turbine-shaped ring in each hexon, and their connecting loops constrict the hexon channel to 14 Å in diameter (Fig. 2i). Because two of the three channel-lining hairpins point inwards, if the pressurized DNA genome should ever reach and propel these hairpins, any hinged outward movement of these hairpins would produce a more constricted channel and thus prevent DNA leakage (Fig. 2h, inset). Six disulfide bonds formed between Cys443 and Cys1231 across neighbouring MCPs (Fig. 2i, inset) further secure the wall of the hexon channel.

When the tower region of penton MCP is superposed with that of hexon MCP, its floor region is tilted approximately 15° towards the capsid centre (Extended Data Fig. 6a, Supplementary Video 3), which is consistent with the more angular geometry of the capsid at five-fold vertices. This hinged movement narrows the penton channel to 5 Å in diameter at the floor level, in contrast to 25 Å at the corresponding position in hexon channel (Extended Data Fig. 6b–d), and separates neighbouring penton MCP tower regions from one another, producing the ‘blossoming’ shape and flexible nature of the penton tower.

Notwithstanding extensive interactions in the MCP tower, three types of network interactions in the MCP floor are the defining features that give rise to the mechanical sturdiness of the KSHV capsid (Fig. 3a–f, Supplementary Videos 4, 5, Extended Data Fig. 7). The type I interaction is intracapsomeric β -augmentation between adjacent MCPs, as exemplified by P2 and P3 MCPs in Fig. 3c: two β -strands in the N-arm of P2 join two β -strands in the E-loop and one β -strand in the dimerization domain of P3 to form a five-stranded β -sheet. Type II and type III interactions are intercapsomeric interactions among two pairs of MCPs, such as P2–P3 and C5–C6 in Fig. 3d. To form a type III interaction, the N-lasso of C5 extends and lashes around the P2 N-arm and P3 E-loop (Fig. 3c, f), which are participants in a type I interaction. In addition, the C5 N-lasso contributes two β -strands to augment the existing five-stranded β -sheet from the type I interaction of P2 and P3, which produces a seven-stranded β -sheet (Fig. 3c). In this regard,

type III interactions build upon and probably strengthen type I interactions. Finally, the short helix in the C5 N-lasso has hydrogen-bond interactions with the P3 helix-hairpin domain and an elbow-like helix–turn–helix structure in the P3 buttress domain, which further secures the C5 N-lasso in place. In turn, this C5 N-lasso serves as a fulcrum for the P3 elbow-like helix–turn–helix structure to support the P3 MCP tower (Fig. 3c). Related by the local two-fold axis, another type III interaction is formed by the P2 N-lasso that lashes around the C5 N-arm and C6 E-loop (Fig. 3f). A type II interaction is formed by two helices from each dimerization domain of P3 or C6 MCP that pair with one another around the local two-fold axis with hydrophobic residues (Fig. 3e). From inside the capsid, this type II interaction appears to act as a pair of ‘snap buttons’ that sit atop the C5 and P2 N-lassoes, prevent their N termini from unwinding and thus lock them in position (Fig. 3d). Therefore, each type II interaction also secures a pair of type III interactions. One can imagine that the force exerted by the DNA genome on the dimerization domains would make the underlying N-lassoes more resistant to unwinding.

As with other herpesviruses^{23,24}, in KSHV the network interactions that surround pentons are different from those of hexons (Fig. 3g, h, Supplementary Video 3). First, the penton MCP has an N terminus that is flexible, rather than one that lashes around the P1–P6 hexon MCP pair as occurs in a canonical type III interaction (Fig. 3g, h). The N-lasso of P6 MCP, which is supposed to lash a pair of penton MCPs, also refolds into a conformation that effectively eliminates its lassoing ability (Fig. 3g). Therefore, a penton neither lashes nor is lashed by adjacent hexons. Second, the dimerization domain of penton MCP adopts a configuration that makes it unable to form type II interactions with the dimerization domain of P1 hexon MCP (Fig. 3h), which renders the P1 dimerization domain flexible. Instead, the refolded dimerization domain of penton MCP contributes one β -strand and the refolded N-lasso of P6 MCP contributes two β -strands—that join the N-arm and E-loop of two penton MCPs—to form a six-stranded β -sheet. This effectively glues the penton together with its surrounding P-hexons (Fig. 3g, h). Finally, the penton MCP has a long straight helix

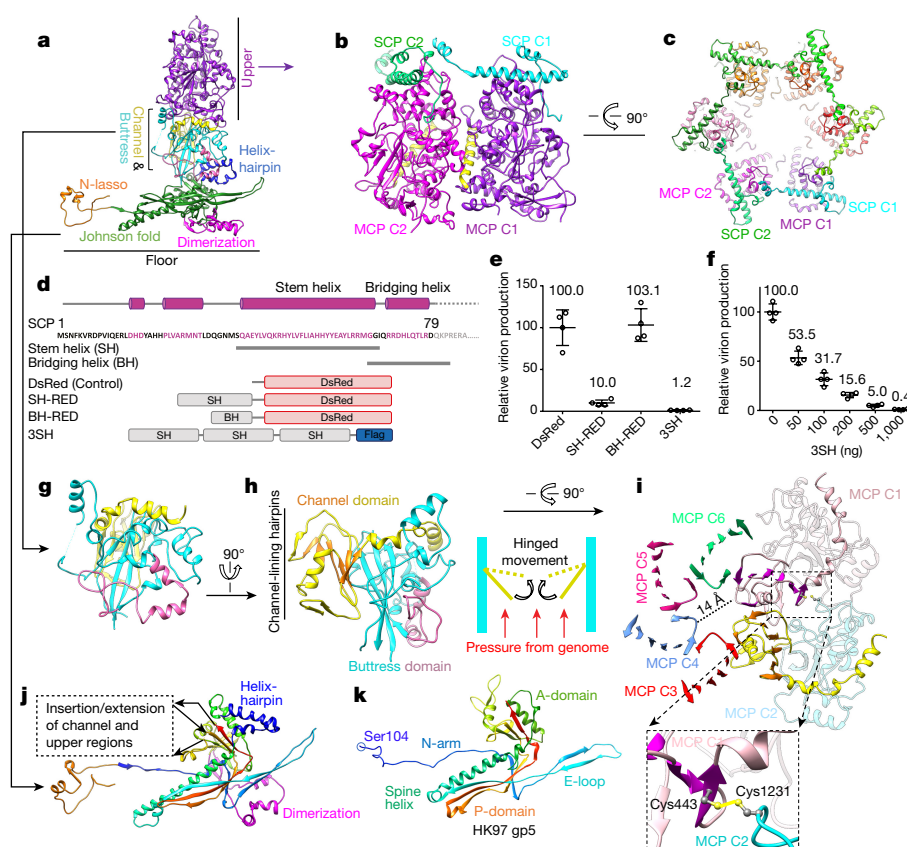


Figure 2 | Structures of MCP and SCP. **a**, Domain organization in a hexon MCP. **b**, **c**, Adjacent MCPs interact around a major helix (yellow in **b**) and are crosslinked by SCPs in hexons. **d–f**, Design of SCP-mimicking polypeptides (**d**) and results of experiment investigating their potential for inhibiting KSHV lytic replication (**e**, **f**). Expression of polypeptides containing the SCP stem helix inhibits KSHV virion production (**e**) in a dose-dependent fashion (**f**). Data are mean \pm s.e.m. ($n = 4$ biologically independent samples). Red, DsRed fragment; SH-RED, fusion protein

of stem helix and DsRed fragment; BH-RED, fusion protein of bridging helix and DsRed fragment. **g–i**, Structure of the MCP channel and buttruss domains. The buttruss domain supports the channel domain (**g**, **h**), and contributes Cys1231 to form a disulfide bond with Cys443 in the channel domain of an adjacent MCP (**i**, inset). **j**, **k**, Structure of the Johnson-fold domain (**j**; with helix-hairpin in dark blue and dimerization in magenta) in the KSHV MCP floor compared with that of bacteriophage HK97 gp5 (**k**).

in its buttruss domain (Fig. 3g) that is refolded from the elbow-like helix–turn–helix structure in hexon MCP (pink in Fig. 2g). This conformational change enables the helix to regain contact with the floor to ‘buttruss’ the penton MCP tower without the P6 N-lasso serving as a fulcrum (Fig. 3c, g).

Of the three types of network interaction, type III—the lashing interaction by the N-lasso—is probably the most important. To investigate its role, we mutated the 60 residues of the N terminal of MCP that encompass the N-lasso and observed the consequences in both viral lytic replication and capsid formation (Fig. 3i–l, Extended Data Fig. 8a, c, e). Removing eight residues from the N terminus of MCP, which retains the N-lasso structure, did not affect viral propagation. By contrast, removing 16 residues from the N terminus of MCP, which renders the N-lasso incomplete, produced a 98.7% decrease in viral titre. Removing 32, 40 or 60 residues, which completely abolishes the N-lasso, reduced virion production to undetectable levels (Fig. 3i, j). Moreover, removing 40 residues of the N terminal (MCP Δ N40) completely abolished capsid assembly (Fig. 3k, l). Together with the above structural analyses, these results establish that the N-lasso of MCP is essential for capsid assembly.

The interaction network of the MCP floor is perforated with holes in the canyons among the towering capsomers, at local or icosahedral three-fold axes. These holes are plugged by triplexes (Fig. 4a, b). Each triplex is a heterotrimer of a Tri1 and two conformers of Tri2—Tri2A and Tri2B (Fig. 4c–f, Supplementary Video 6). Each Tri2 monomer has a β -sheet-rich trunk domain (amino acids 2–141, 283–305) and

a helix-rich ‘embracing’ arm domain (142–282). Tri2A and Tri2B embrace one another to form a dimer (Fig. 4i), which is accommodated by conformational changes in their embracing arm domains (Fig. 4g, Supplementary Video 7). This embracing interaction involves large patches of hydrophobic residues (Extended Data Fig. 9a–d) and two disulfide bonds that are formed between Cys212 of one conformer and Cys222 of the other (Fig. 4i). Supporting the embracing Tri2A–Tri2B dimer is the ‘third-wheel’ domain of the Tri1 monomer, which is structurally similar to the trunk domain of Tri2 (Fig. 4h). The N-terminal region of Tri1 penetrates the capsid floor and folds into a tripod-shaped daisy-chain of helices (the N-anchor) that binds to three hydrophobic grooves formed by the spine helix and its associated β -sheet in the Johnson-fold domains of three adjacent MCPs (Fig. 4j, Extended Data Fig. 9e–g). Thus, the N-anchor of Tri1 anchors the entire triplex to the capsid floor; this plugs the hole in the MCP floor and rivets the capsid shell in place. The N-anchor of Tri1 in peripentonal triplex Ta refolds in the middle arm of the tripod to accommodate the binding of a helix from the refolded dimerization domain of a penton MCP, and thus probably stabilizes the penton region (Extended Data Fig. 10).

To investigate the function of the Tri1 N-anchor, we carried out serial-truncation mutageneses of the 65 residues of the Tri1 N terminal (Fig. 4j–m, Extended Data Fig. 8b, d, f). Truncating 20, 40 and 65 residues of the Tri1 N terminal, which removes one, two and three tripod arms, reduced virion production by 88.4%, 99.7% and to an undetectable level, respectively (Fig. 4j, k). Furthermore, we observed

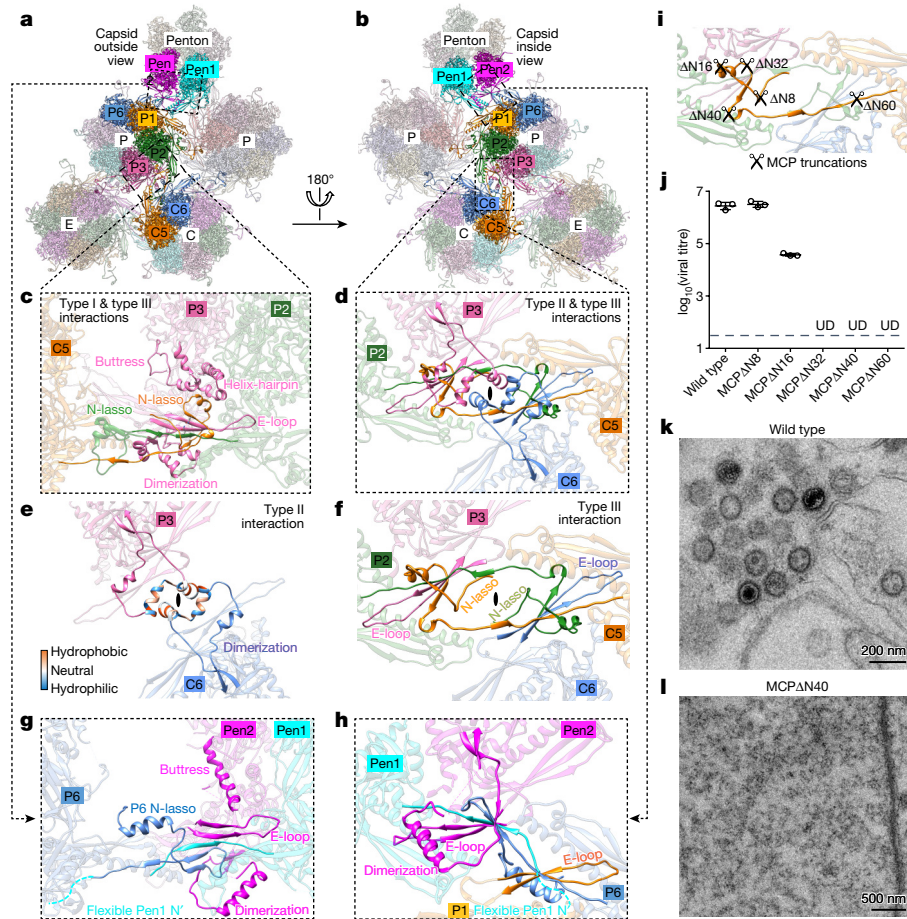


Figure 3 | Network interactions in the MCP floor and function of MCP N-lasso. **a, b**, Part of the MCP network viewed from outside (**a**) or inside (**b**) the capsid. Pen1 and Pen2 are two of the five MCP subunits in the penton that have the same structure owing to five-fold symmetry of the penton. **c–f**, Three types of network interaction among hexon MCPs; **e** and **f** represent decomposition of structures in **d**. Type I interactions are intracapsomeric augmentations of β -strands from adjacent MCPs (P2 and P3 in **c**). Type II and type III interactions are intercapsomeric interactions among two pairs of MCPs (P2–P3 and C5–C6 in **d**), diagonally across the local two-fold axis (black ovals in **d–f**). Type III interactions build upon and fortify type I interactions (C5 N-lasso in **c**). The two dimerization domains that are joined in a type II interaction (**e**) also sit on top of (from the perspective of inside of the capsid) a pair of type III interactions (**d**)

the accumulation of empty capsid-like particles in the nuclei of cells that lytically replicated the Tri1 Δ N65 mutant virus (Fig. 4l, m), in contrast to the observation with MCP Δ N40 mutant virus, in which lack of virion production was probably due to complete abolition of capsid assembly (Fig. 3k, l). Although triplex has been proposed to incorporate three MCP subunits into an assembly unit for procapsid formation²⁵, our mutagenesis results indicate that the Tri1 N-anchor is not required for triplex incorporation during capsid assembly, though it is essential for DNA containment in the C-capsids.

The hotspots of interactions revealed in the above structural and mutational analyses can be targeted for designing capsid assembly inhibitors. Current therapies against herpesvirus infections are based on nucleoside analogues that target viral genome replication, a process that bears some similarity to cellular DNA replication. These therapies therefore suffer from cytotoxicity and drug resistance²⁶, which generates a need for potent non-nucleosidic antiviral compounds. Compounds that target capsid formation or stability are a potential option and have been actively pursued in research focusing on HIV and hepatitis B^{27–30}; owing to the absence of atomic structures for

and prevent the two N-lassoes (C5 and P2 N-lassoes in **d** and **f**) from unwinding. **g, h**, Interactions between penton MCPs and the P1 and P6 hexon MCPs are different from the canonical hexon MCP network interactions (shown in **c** and **d**). An elbow-like helix–turn–helix structure in the buttress domain of hexon MCP is folded into a single long helix in penton MCP (see **c** and **g**). **i**, Design of serial-truncation mutageneses in the MCP N-lasso region. **j**, Comparison of viral titres of wild type and MCP-truncated mutants. Data are mean \pm s.e.m. ($n = 3$ biologically independent samples). UD, undetectable. **k, l**, Transmission electron microscopy images of ultrathin sections of cells replicating the wild-type virus (**k**) or the MCP Δ N40 mutant (**l**). Experiments were repeated independently twice with similar results.

herpesvirus capsids, similar strategies have been hindered in herpes research. Our structure reveals an interaction hotspot that seems to offer a promising target for developing antiviral agents: the hydrophobic groove on the MCP surface, into which the SCP stem helix binds. We reason that a short polypeptide that mimics the SCP stem helix would compete with wild-type SCP for binding to MCP, interrupt the crosslinking and stabilizing effect of the SCP, and thereby inhibit KSHV lytic replication. To test this strategy, we constructed three polypeptides that contained the stem helix, the bridging helix or three tandem repeats of the stem helix (3SH) (Fig. 2d). When expressed in cells, 3SH polypeptides interacted with the MCPud in a similar way to wild-type SCP (Extended Data Fig. 5b, c). Expression of stem helix or 3SH polypeptides in KSHV-replicating cells reduced virion production by 90% and 98.8%, respectively; expression of bridging helix polypeptides did not affect virion production (Fig. 2e). Further experiments with 3SH polypeptides indicated that its inhibition of virion production is dose dependent (Fig. 2f). Therefore, the SCP stem helix is a starting point or even a possible lead compound for the development of drugs targeting KSHV. If this SCP-mimicking polypeptide proves to be useful in

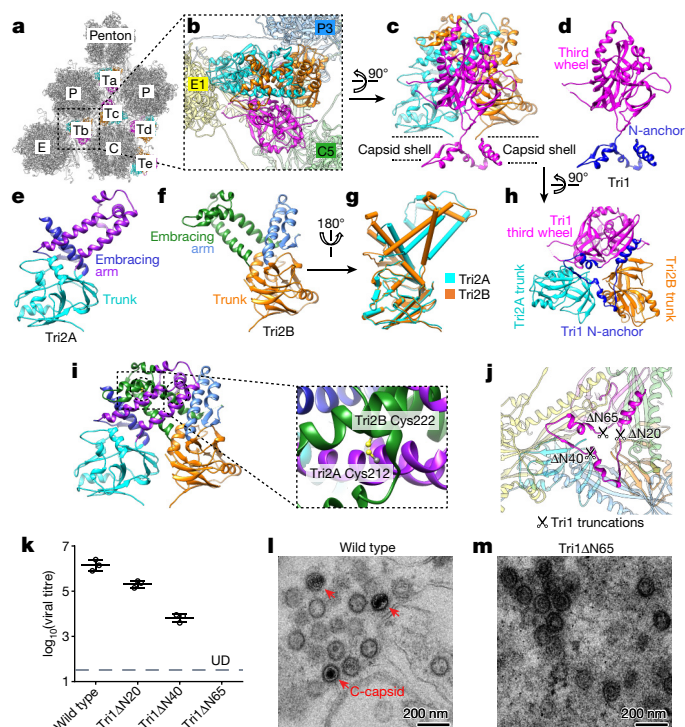


Figure 4 | Structure of triplex and function of Tri1 N-anchor.

a, Distribution of triplexes in the MCP network. **b**, Enlarged view of a triplex Tb from outside the capsid. **c–f**, Detailed structures of triplex Tb (**c**) and its components Tri1 (**d**), Tri2A (**e**) and Tri2B (**f**). **g**, Superposition of Tri2A and Tri2B. **h**, Triplex Tb viewed from inside the capsid showing similar structures among the Tri2A and Tri2B trunk domains and the Tri1 third-wheel domain. **i**, Tri2A and Tri2B form a dimer with their embracing arms. The dotted circles denote two disulfide bonds between Tri2A and Tri2B (shown expanded in the inset). **j**, Triplex Tb viewed from inside the capsid, showing that it anchors to the capsid floor by the tripod-shaped Tri1 N-anchor. **k**, Comparison of viral titres of wild type and Tri1-truncated mutants. Data are mean \pm s.e.m. ($n = 3$ biologically independent samples). **l, m**, Transmission electron microscopy images of ultrathin sections of cells replicating the wild-type virus (**l**) or the Tri1 Δ N65 mutant (**m**). Experiments were repeated independently twice with similar results.

designing effective small molecule inhibitors against KSHV infection, many other interaction hotspots revealed in our atomic structure may also offer resources for further exploration.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 April; accepted 6 December 2017.

Published online 17 January 2018.

- Chang, Y. *et al.* Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266**, 1865–1869 (1994).
- Moore, P. S. & Chang, Y. Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma in patients with and those without HIV infection. *N. Engl. J. Med.* **332**, 1181–1185 (1995).
- Boshoff, C. & Weiss, R. AIDS-related malignancies. *Nat. Rev. Cancer* **2**, 373–382 (2002).
- Chatlynne, L. G. & Ablashi, D. V. Seroepidemiology of Kaposi's sarcoma-associated herpesvirus (KSHV). *Semin. Cancer Biol.* **9**, 175–185 (1999).
- Bauer, D. W., Huffman, J. B., Homa, F. L. & Evilevitch, A. Herpes virus genome, the pressure is on. *J. Am. Chem. Soc.* **135**, 11216–11221 (2013).
- Perkins, E. M. *et al.* Small capsid protein pORF65 is essential for assembly of Kaposi's sarcoma-associated herpesvirus capsids. *J. Virol.* **82**, 7201–7211 (2008).
- Henson, B. W., Perkins, E. M., Cothran, J. E. & Desai, P. Self-assembly of Epstein–Barr virus capsids. *J. Virol.* **83**, 3877–3890 (2009).
- Dai, X. *et al.* CryoEM and mutagenesis reveal that the smallest capsid protein cements and stabilizes Kaposi's sarcoma-associated herpesvirus capsid. *Proc. Natl Acad. Sci. USA* **112**, E649–E656 (2015).

- Dai, X. *et al.* The smallest capsid protein mediates binding of the essential tegument protein pp150 to stabilize DNA-containing capsids in human cytomegalovirus. *PLoS Pathog.* **9**, e1003525 (2013).
- Zhang, X. & Zhou, Z. H. Limiting factors in atomic resolution cryo electron microscopy: no simple tricks. *J. Struct. Biol.* **175**, 253–263 (2011).
- Zhou, Z. H. *et al.* Four levels of hierarchical organization, including noncovalent chainmail, brace the mature tumor herpesvirus capsid against pressurization. *Structure* **22**, 1385–1398 (2014).
- Wikoff, W. R. *et al.* Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129–2133 (2000).
- Fokine, A. *et al.* Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proc. Natl Acad. Sci. USA* **102**, 7163–7168 (2005).
- Zhang, X. *et al.* A new topology of the HK97-like fold revealed in Bordetella bacteriophage by cryoEM at 3.5 Å resolution. *eLife* **2**, e01299 (2013).
- Lander, G. C. *et al.* Bacteriophage lambda stabilization by auxiliary protein gpD: timing, location, and mechanism of attachment determined by cryo-EM. *Structure* **16**, 1399–1406 (2008).
- Parent, K. N. *et al.* P22 coat protein structures reveal a novel mechanism for capsid maturation: stability without auxiliary proteins or chemical crosslinks. *Structure* **18**, 390–401 (2010).
- Baker, M. L., Jiang, W., Rixon, F. J. & Chiu, W. Common ancestry of herpesviruses and tailed DNA bacteriophages. *J. Virol.* **79**, 14967–14970 (2005).
- Akita, F. *et al.* The crystal structure of a virus-like particle from the hyperthermophilic archaeon *Pyrococcus furiosus* provides insight into the evolution of viruses. *J. Mol. Biol.* **368**, 1469–1483 (2007).
- Sutter, M. *et al.* Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol.* **15**, 939–947 (2008).
- Bowman, B. R., Baker, M. L., Rixon, F. J., Chiu, W. & Quiocho, F. A. Structure of the herpesvirus major capsid protein. *EMBO J.* **22**, 757–765 (2003).
- Dai, X., Gong, D., Wu, T. T., Sun, R. & Zhou, Z. H. Organization of capsid-associated tegument components in Kaposi's sarcoma-associated herpesvirus. *J. Virol.* **88**, 12694–12702 (2014).
- Zhou, Z. H. *et al.* Assembly of VP26 in herpes simplex virus-1 inferred from structures of wild-type and recombinant capsids. *Nat. Struct. Mol. Biol.* **2**, 1026–1030 (1995).
- Huet, A. *et al.* Extensive subunit contacts underpin herpesvirus capsid stability and interior-to-exterior allostery. *Nat. Struct. Mol. Biol.* **23**, 531–539 (2016).
- Yu, X., Jih, J., Jiang, J. & Zhou, Z. H. Atomic structure of the human cytomegalovirus capsid with its securing tegument layer of pp150. *Science* **356**, eaam6892 (2017).
- Aksyuk, A. A. *et al.* Subassemblies and asymmetry in assembly of herpes simplex virus procapsid. *MBio* **6**, e01525–15 (2015).
- Gilbert, C., Bestman-Smith, J. & Boivin, G. Resistance of herpesviruses to antiviral drugs: clinical impacts and molecular mechanisms. *Drug Resist. Updat.* **5**, 88–114 (2002).
- Ternois, F., Sticht, J., Duquerroy, S., Kräusslich, H. G. & Rey, F. A. The HIV-1 capsid protein C-terminal domain in complex with a virus assembly inhibitor. *Nat. Struct. Mol. Biol.* **12**, 678–682 (2005).
- Sticht, J. *et al.* A peptide inhibitor of HIV-1 assembly *in vitro*. *Nat. Struct. Mol. Biol.* **12**, 671–677 (2005).
- Deres, K. *et al.* Inhibition of hepatitis B virus replication by drug-induced depletion of nucleocapsids. *Science* **299**, 893–896 (2003).
- Zlotnick, A., Ceres, P., Singh, S. & Johnson, J. M. A small molecule inhibits and misdirects assembly of hepatitis B virus capsids. *J. Virol.* **76**, 4848–4854 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements This project was supported in part by grants from the National Institutes of Health (NIH) (DE025567, GM071940, AI094386, CA091791 and CA177322) and indirectly through a Clinical and Translational Science Institute core voucher award (UL1TR000124) from UCLA's National Center for Advancing Translational Science. We acknowledge the use of instruments at the Electron Imaging Center for Nanomachines supported by UCLA and by instrumentation grants from the NIH (1S10OD018111, 1U24GM116792) and NSF (DBI-1338135, DMR-1548924).

Author Contributions Z.H.Z., X.D., D.G. and R.S. designed the project; Z.H.Z., R.S. and T.-T.W. supervised research; D.G. and X.D. prepared the samples; X.D. acquired cryo-EM data and determined the structure; X.D., H.L. and J.J. built atomic models; D.G. performed functional studies; X.D., D.G., Z.H.Z., R.S. and T.-T.W. interpreted the results; Z.H.Z., X.D. and D.G. wrote the paper; R.S. revised the paper; and all authors reviewed the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to R.S. (RSun@mednet.ucla.edu, for virology and functional analyses) and Z.H.Z. (Hong.Zhou@UCLA.edu, for cryo-EM and atomic modelling).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Sample preparation. Culture and purification of KSHV virions followed a previously described procedure^{8,21}. In brief, KSHV lytic replication was induced in an iSLK-puro cell line that harbours the KSHV-BAC16 plasmid^{31,32}. The tissue culture supernatant was collected and KSHV virions were pelleted by ultracentrifugation and purified by sucrose density gradient^{9,33}. An aliquot of 2.5 µl purified virion sample was applied onto a Quantifoil R2/1 Cu grid, manually blotted with filter paper and plunge-frozen in liquid ethane. The large size (over 200 nm) of the KSHV virion and frequent contamination of even larger (up to 500 nm) virus-like vesicles in the density-gradient-purified sample (Extended Data Fig. 1a) presented a major challenge to obtaining an optimal cryo-EM grid that had both fully embedded particles and adequately thin ice for high-resolution cryo-EM imaging. Moreover, in order to collect enough data for high-resolution structure determination, the sample must have a 'perfect' concentration so that at least a few particles can be imaged within each micrograph but not be over-concentrated, which results in overlapping particles and thick ice. We experimented with more than 30 rounds of virus culture and purification, and screened hundreds of cryo-EM grids to obtain just two that were deemed adequate for high-resolution cryo-EM imaging. The best cryo-EM grids for KSHV virions we obtained were prepared by manual blotting with the filter paper parallel to the grid surface rather than with the commonly used Vitrobot, in which the filter paper blots the grid at an angle and the excess sample solution flows to the filter paper from one edge of the grid.

Cryo-EM imaging and data preprocessing. Cryo-EM images were recorded in an FEI Titan Krios cryo-electron microscope. We collected 8,007 movies using Legikon³⁴ with a Gatan K2 Summit direct electron detector in super-resolution mode. Each movie consists of 26 frames, each of which has an exposure time of 500 ms. A dose rate of approximately 8 electrons per hardware pixel per second was set for exposure to facilitate electron counting in the camera. The pixel size in K2 super-resolution images at 14,000× nominal magnification was calibrated to be 1.03 Å per pixel. Thus, the accumulated dose for each movie was approximately 25 e⁻ per Å² on the sample. All 26 frames in each movie were aligned and averaged for drift correction³⁵. Defocus values were determined with CTFFIND3³⁶ to be in the range of -1 µm to -3 µm. Particles were picked with Etha³⁷ and then manually screened with the boxer program in EMAN³⁸ to keep only well-separated and artefact-free particles. A total of 44,343 particle images were boxed out from the micrographs with EMAN.

3D structure determination. The enormous size of 1,440×1,440 pixels for individual particle image is too large or requires an unrealistic amount of computational resources to process with many popular cryo-EM software packages. To process the data, we modified and rebuilt our common-line based refinement program package (IMIRS^{39,40}) and GPU-implemented reconstruction program (eLite3D⁴¹), to expand data processing capacities. The original images were also binned 8×, 4× or 2× stepwise to speed up data processing.

Initial orientation and centre parameters for each particle were determined using data points in the very low resolution range (500–30 Å), without contrast transfer function correction. A phase-residue value was reported for each particle that directly measured the agreement of the cross common-lines between the cryo-EM image and the projected template. High phase-residue values typically result from incorrect identifications of the cross common-lines (that is, incorrect orientation assignment for that particle) or from low-quality particles (such as broken or contaminated particles). By sorting and grouping the particles with phase-residue values in steps of 0.4 and plotting the number of particles in each group, we found that there were two well-separated peaks of phase-residue distribution in the dataset (Extended Data Fig. 1b). The number of particles in the high phase-residue peak constituted approximately 35% of the dataset. We checked a previous dataset (X.D. *et al.*, unpublished data) of the KSHV virion that was recorded on photographic film and found that the high phase-residue peak constituted an even higher percentage (55%) of that dataset (Extended Data Fig. 1c), probably owing to a lower contrast of the film dataset in comparison with our current K2 super-resolution dataset. Because only low-resolution (and thus relatively less noisy) data points were used in the initial centre and orientation search, the high phase-residue particles should be treated as 'bad particles' and the low phase-residue ones as 'good particles'. The relatively high percentage of bad particles in the KSHV dataset compared to that in viral datasets we have previously worked on (for example, refs 42, 43) might be caused by relatively thick ice of the KSHV cryo-EM grid and, more probably, by the thick and pleomorphic tegument layer in the KSHV virion, which not only deteriorates the contrast of the cryo-EM images but also acts as a contaminant to the capsid projection, which interferes with centre and orientation searches for the capsid. As the refinement gradually

proceeds to include more and more high-resolution data points, the phase residue of all particles would shift towards the 90° maximum. Conceivably, the phase residue of a good particle would shift faster than that of a bad particle, so the two well-separated peaks observed at the initial stage would gradually overlap with one another. Therefore, at a late stage of the refinement procedure, it would be difficult or even impossible to use a single phase-residue cutoff to select most of the good particles for 3D reconstruction. To prevent the bad particles from 'contaminating' the final reconstruction, we adopted a strategy in which bad particles were sorted out and thrown away after the initial step of centre and orientation search, and before the refinement procedure. A total of 29,100 high-quality particles in the first peak of phase-residue distribution, constituting approximately 65% of the entire dataset, were selected, divided into two random halves and separately subjected to iterative refinement. At the end of each iteration, a phase-residue cutoff was set to select the top approximately 85% of particles for reconstruction (Extended Data Fig. 1d).

After convergence of refinement for both halves, a Fourier shell correlation (FSC) curve was calculated and the resolution was determined to be 4.2 Å on the basis of the gold-standard FSC = 0.143 criterion⁴⁴. Then, the two halves of the dataset were combined and a total of 25,315 particles were used to calculate the final density map. A B-factor of -200 Å² was applied to sharpen the density map for local building and structure analysis.

Local averaging. There are 15 copies of hexon MCP or SCP (C1–C6, E1–E3 and P1–P6) and five copies of triplex (Ta–Te) in the asymmetric unit of the KSHV capsid. The quasi-equivalent copies of each kind are structurally similar and thus can be averaged to boost the signal-to-noise ratio of the density map, and to facilitate backbone tracing during atomic model building. By segmenting out cuboid density blocks encompassing each copy of MCP, SCP or triplex, and fitting them to each other in Chimera⁴⁵, we identified that densities of MCP or SCP C1–C6 or triplexes Tb–Te have the best quality and are structurally least different among their kind. These aligned density blocks in each kind were averaged with the vop command in Chimera.

Atomic model building. The local-averaged MCP, SCP or triplex density maps were fitted into the original map at one of the quasi-equivalent positions and both were used for *ab initio* modelling in Coot⁴⁶. Generally, the local-averaged map has much improved main-chain densities that help to resolve ambiguities of backbone tracing; whereas the original map has better side chain densities that provide landmarks for amino acid registration. The crystallographic model of HSV-1 MCPud (RSCB Protein Data Bank (PDB) code: 1NO7)²⁰ fits into our KSHV density map and was therefore referred to when tracing backbone for the KSHV MCPud.

The first atomic model of MCP, SCP or triplex was refined using the real space refinement utility in Phenix⁴⁷ and then manually checked in Coot. This procedure was repeated several times until a satisfactory model was obtained. These first models of each kind were fitted into the original density map at other quasi-equivalent positions and manually adjusted in Coot for structural variations. For penton MCP, a majority of the tower region (including upper, channel and buttress domains) is noisy and broken in the sharpened map, and is therefore difficult to model *ab initio*. However, by low-pass filtering the density map to 6 Å resolution (so that the noisy penton tower could be visualized), and fitting a hexon MCP model into the penton density, we found that the tower regions of penton MCP and hexon MCP are structurally almost identical. In the floor region, where the two structures differ substantially, the penton density shows adequate quality for *ab initio* modelling. Therefore, the tower region of hexon MCP model is chimerized with the *ab initio* model of the penton MCP floor region to synthesize the full model of penton MCP for presentation purposes.

In summary, we built atomic models for a total of 46 unique conformers of the four capsid proteins: 15 hexon MCP (amino acids 1–1,141, 1,164–1,376), 1 penton MCP (*ab initio* modelling of amino acids 47–409, 1,261–1,293 except for small flexible regions, and fitting of the remaining with hexon MCP model), 15 hexon SCP (amino acids 2–79), 5 Tri1 (amino acids 4–213, 217–331) and 10 Tri2 (amino acids 2–163, 174–305 of Tri2A; 2–196, 201–305 of Tri2B). As the last step, the refined atomic models of all individual conformers were combined and refined together in Phenix to resolve intermolecular clashes at the interface. Though not resolved at high resolution, the densities of five capsid-associated tegument complexes, which crown each penton, are visible when the map is low-pass filtered to approximately 6 Å resolution, suggesting their relative flexibility/low occupancy²¹. This resolution was not sufficient to support model building for the capsid-associated tegument complexes.

Construction of MCP-truncated or Tri1-truncated KSHV mutants. The KSHV-BAC16 plasmid was modified according to a previously described method^{32,48}. In brief, DNA fragments of KSHV ORF25 (MCP) or ORF62 (Tri1) with truncation of the defined amino acids were used to replace the wild-type sequence in the KSHV BAC by homogeneous recombination in *Escherichia coli*. Restriction patterns of the mutated KSHV BACs were verified by comparing them to that of the wild type

to ensure overall genome integrity without gross changes. Fragments with the mutations in the middle were amplified from the plasmids by PCR, and sequenced to confirm that all mutations were correct. All mutant KSHV BACs were introduced into iSLK-puro cells, followed by selection with 1,200 $\mu\text{g ml}^{-1}$ hygromycin B, 1 $\mu\text{g ml}^{-1}$ puromycin and 250 $\mu\text{g ml}^{-1}$ G418 for one month to generate cell lines latently infected by a specific KSHV mutant virus.

Titration of infectious KSHV virions. To determine the concentration of infectious KSHV virions released from iSLK-puro cells harbouring a wild-type, MCP-truncated or Tri1-truncated KSHV genome, cells were treated with 1 mM sodium butyrate plus 1 $\mu\text{g ml}^{-1}$ doxycycline for three days to induce KSHV lytic replication. Then, the supernatants were collected, centrifuged at 10,000g for 10 minutes at 4 °C to remove cellular debris, serially diluted in DMEM with 10% FBS and used to infect 293T cells in 96-well plates by spinoculation (3,000g for 1 h at 30 °C). Three days after infection, GFP-positive cell clusters containing two or more cells were counted under a fluorescence microscope to determine the titre of KSHV virions. Infectious units are expressed as the number of GFP-positive cell clusters in each well at a specific dilution of the viral stock.

Measuring viral DNA replication and RNA expression by real-time PCR. Total DNA was isolated from cells replicating KSHV, and viral genome copy numbers were determined by real-time PCR using primers for ORF59, a gene essential for viral genome replication⁴⁹. Total RNA was extracted from cells with a Purelink RNA mini kit (Thermo Fisher Scientific), treated with DNase I and reverse transcribed using SuperScript III reverse transcriptase (Thermo Fisher Scientific) and random hexamers. Real-time PCR was then performed with the following primers to detect the corresponding DNA or transcripts. *GAPDH*: 5'-TGCACCACTGCTTAGC-3' and 5'-GGCATGGACTGTGGTCATGAC-3'; *RTA*: 5'-CACAAAATGGCGCAAGATGA-3' and 5'-TGGTAGAGTTGGGCCTTCAGTT-3'; *ORF59*: 5'-TTGGCACTCCAACGAAATATTAGAA-3' and 5'-CGGGAACCTTTTGCAGAAGA-3'; *ORF57*: 5'-TGGACATTATGAAGGGCATCTA-3' and 5'-CGGGTTTCGGACAATTGCT-3'; *ORF52*: 5'-CTTACGATGGAGACCTAACCG-3' and 5'-ATCCAGTGCTTTCCGAAG-3'; *ORF25* (MCP): 5'-CGTATCCCCTGTTCTGCTATG-3' and 5'-TTTCCCGAGTTGACCCAG-3'; *ORF62* (Tri1): 5'-TCGTGGTTTATCTCCGTGTG-3' and 5'-CAGCTGAATATATCTTGTCGGG-3'.

Western blotting and antibodies. Proteins in SDS-PAGE sample buffer were heated at 95 °C, resolved by SDS-PAGE and then transferred onto PVDF membrane. Proteins were detected with antibodies against KSHV RTA (polyclonal antibody produced in rabbit), SCP (polyclonal antibody produced in rabbit⁵⁰), Tri1 (Thermo Fisher Scientific) or actin (Abcam). HRP-conjugated secondary antibodies were used and detection was performed with ECL (Amersham).

Plastic embedding, ultrathin sectioning and conventional transmission electron microscopy. Cells induced for lytic replication of the wild-type, MCP-truncated or Tri1-truncated KSHV were collected after trypsin treatment and subjected to plastic embedding and transmission electron microscopy as previously described⁸. In brief, cells were washed with PBS, fixed with 2% glutaraldehyde in PBS for 1 h, post-fixed in 1% OsO₄ for 1 h, stained en bloc in 2% uranyl acetate for 1 h, dehydrated in an ascending ethanol series and embedded using the Ultra Bed Low Viscosity Epoxy Kit (Electron Microscopy Sciences). Approximately 75-nm-thick sections were stained with saturated aqueous uranyl acetate and lead citrate, and examined with an FEI Tecnai F20 electron microscope.

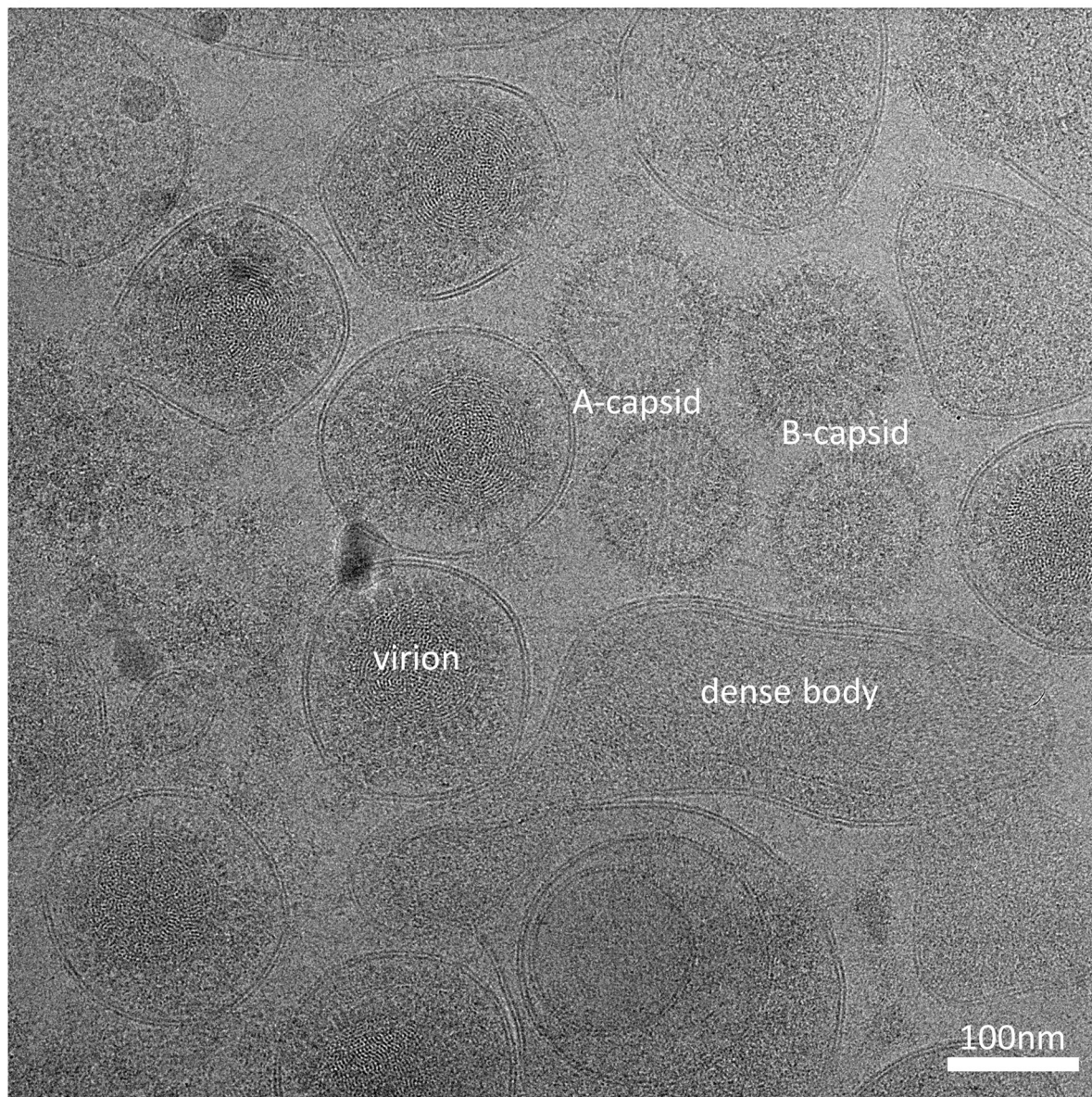
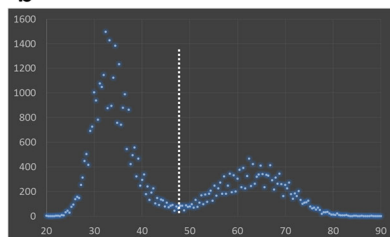
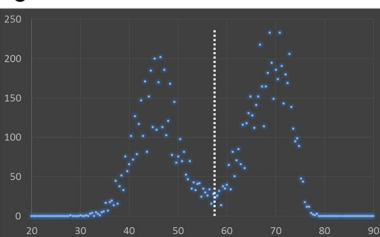
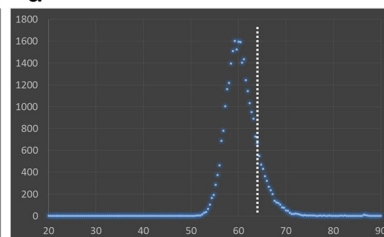
Co-immunoprecipitation assay to detect MCP-SCP binding. Co-immunoprecipitation experiments were performed as previously described⁵¹. In brief, 293T cells were transfected with SCP-Flag and Myc-MCPud (amino acids 478–1,033) expression plasmids and lysed at two days post-transfection with RIPA buffer (50 mM Tris-HCl pH 7.4, 0.5% NP-40, 150 mM NaCl, 1 mM EDTA and protease inhibitors). After pre-clearance with protein G-sepharose beads (GE Health) for 1 h at 4 °C, cell lysates were incubated with 1 μg anti-Myc antibody for at least 4 h at 4 °C with constant agitation. Then, protein complexes were collected by incubating with protein G-sepharose beads for 1 h, and further washed five times with RIPA buffer. Proteins were eluted from beads with 60 μl SDS-PAGE sample buffer and subjected to western blotting analysis using anti-Flag and anti-Myc antibodies.

Construction and test of SCP-mimicking polypeptides. The coding region for the SCP stem helix (amino acids 37–68) or bridging helix (amino acids 66–81) was amplified from KSHV-BAC16 plasmid by PCR, and the coding region for DsRed was amplified from pCMV-DsRed-Express plasmid (Clontech) by PCR. Stem helix or bridging helix fragments were linked to DsRed fragment by PCR

reaction, and then cloned into the EcoRI site of pCMV-noHA vector⁵¹ to generate SH-RED and BH-RED polypeptide expression plasmids, respectively. The DNA sequence for three tandem repeats of stem helix linked with Ser-Gln-Pro residues and C-terminal Flag-tag was synthesized *de novo* (Integrated DNA Technologies), and then cloned into the EcoRI site of pCMV-noHA vector to generate the 3SH polypeptide expression plasmid. Sequences of the PCR-amplified or synthesized fragments and their correct insertion in the plasmid were verified by sequencing. KSHV latently infected 293T cells were transfected with these polypeptide expression plasmids or the pCMV-DsRed-Express vector as control. At 16 h post-transfection, cells were treated with 0.5 mM sodium butyrate plus 25 ng ml⁻¹ TPA to induce KSHV lytic replication. Three days later, the supernatants were collected for virion titration.

Data availability. The cryo-EM density map and the atomic models have been deposited in the Electron Microscopy Data Bank and RCSB Protein Data Bank under accession numbers EMD-7047 and 6B43, respectively. All other data are available from the corresponding authors upon reasonable request.

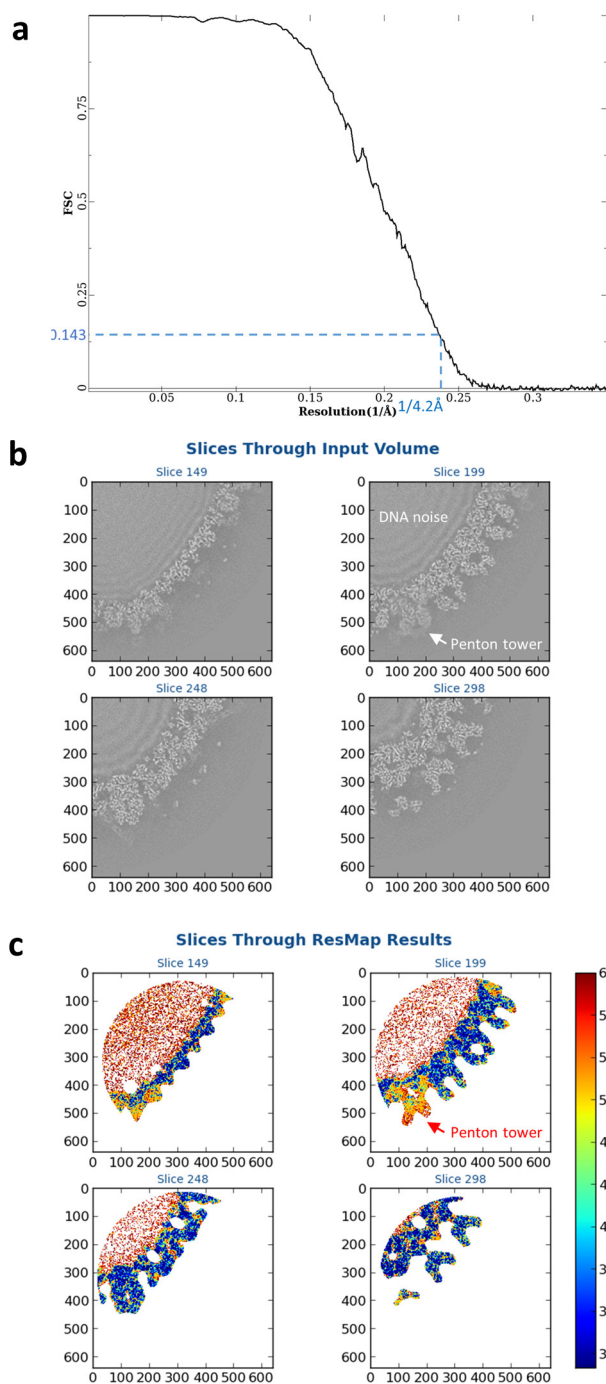
- Myoung, J. & Ganem, D. Generation of a doxycycline-inducible KSHV producer cell line of endothelial origin: maintenance of tight latency with efficient reactivation upon induction. *J. Virol. Methods* **174**, 12–21 (2011).
- Brulois, K. F. *et al.* Construction and manipulation of a new Kaposi's sarcoma-associated herpesvirus bacterial artificial chromosome clone. *J. Virol.* **86**, 9708–9720 (2012).
- Dai, X. & Zhou, Z. H. Purification of Herpesvirus virions and capsids. *Bio Protoc.* **4**, e1193 (2014).
- Suloway, C. *et al.* Automated molecular microscopy: the new Legimon system. *J. Struct. Biol.* **151**, 41–60 (2005).
- Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
- Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
- Kivioja, T., Rantanen, J., Verkhovsky, A., Ukkonen, E. & Bamford, D. Local average intensity-based method for identifying spherical particles in electron micrographs. *J. Struct. Biol.* **131**, 126–134 (2000).
- Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
- Liang, Y., Ke, E. Y. & Zhou, Z. H. IMIRIS: a high-resolution 3D reconstruction package integrated with a relational image database. *J. Struct. Biol.* **137**, 292–304 (2002).
- Liu, H. *et al.* Symmetry-adapted spherical harmonics method for high-resolution 3D single-particle reconstructions. *J. Struct. Biol.* **161**, 64–73 (2008).
- Zhang, X., Zhang, X. & Zhou, Z. H. Low cost, high performance GPU computing solution for atomic resolution cryoEM single-particle reconstruction. *J. Struct. Biol.* **172**, 400–406 (2010).
- Zhang, X., Jin, L., Fang, Q., Hui, W. H. & Zhou, Z. H. 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell* **141**, 472–482 (2010).
- Yu, X., Jiang, J., Sun, J. & Zhou, Z. H. A putative ATPase mediates RNA transcription and capping in a dsRNA virus. *Elife* **4**, e07901 (2015).
- Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
- Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
- Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
- Gong, D. *et al.* Kaposi's sarcoma-associated herpesvirus ORF18 and ORF30 are essential for late gene expression during lytic replication. *J. Virol.* **88**, 11369–11382 (2014).
- Chan, S. R. & Chandran, B. Characterization of human herpesvirus 8 ORF59 protein (PF-8) and mapping of the processivity and viral DNA polymerase-interacting domains. *J. Virol.* **74**, 10920–10929 (2000).
- Lin, S. F. *et al.* Identification, expression, and immunogenicity of Kaposi's sarcoma-associated herpesvirus-encoded small viral capsid antigen. *J. Virol.* **71**, 3069–3076 (1997).
- Gong, D. *et al.* A herpesvirus protein selectively inhibits cellular mRNA nuclear export. *Cell Host Microbe* **20**, 642–653 (2016).
- Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
- Helgstrand, C. *et al.* The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution. *J. Mol. Biol.* **334**, 885–899 (2003).

a**b****c****d**

Extended Data Figure 1 | Cryo-EM imaging of KSHV particles and data processing strategy to minimize interference of the tegument layer.

a, A cryo-EM micrograph. Images recorded with a Gatan K2 Summit direct electron detector show largely intact KSHV virions in our sample preparation. Naked capsids were only occasionally observed. The defocus value of this micrograph was $-1.3\ \mu\text{m}$. **b**, Plot of phase-residue value distribution of particles after initial determination of orientation and centre parameters. Particles in the second peak with high phase-residue values are regarded as bad particles, for which the parameters were not correctly determined owing to the interference of the thick, pleomorphic tegument layer or the low quality of the particle. These particles were

discarded and not included for following refinement to avoid their contaminating the reconstruction. **c**, Plot of phase-residue value distribution of our previous film dataset (X.D. *et al.*, unpublished data) showing even more bad particles than those recorded in **b**, probably owing to decreased contrast of the film dataset compared to that of our current K2 dataset (**b**). **d**, Plot of phase-residue value distribution of particles after the final round of parameter refinement. There is only one peak representing the good particles, because the bad particles were discarded at the beginning of the refinement procedure. Moreover, only the top 85% of these good particles were selected for reconstruction. White lines denote the phase-residue value cutoff to select particles for refinement (**b**, **c**) or reconstruction (**d**).

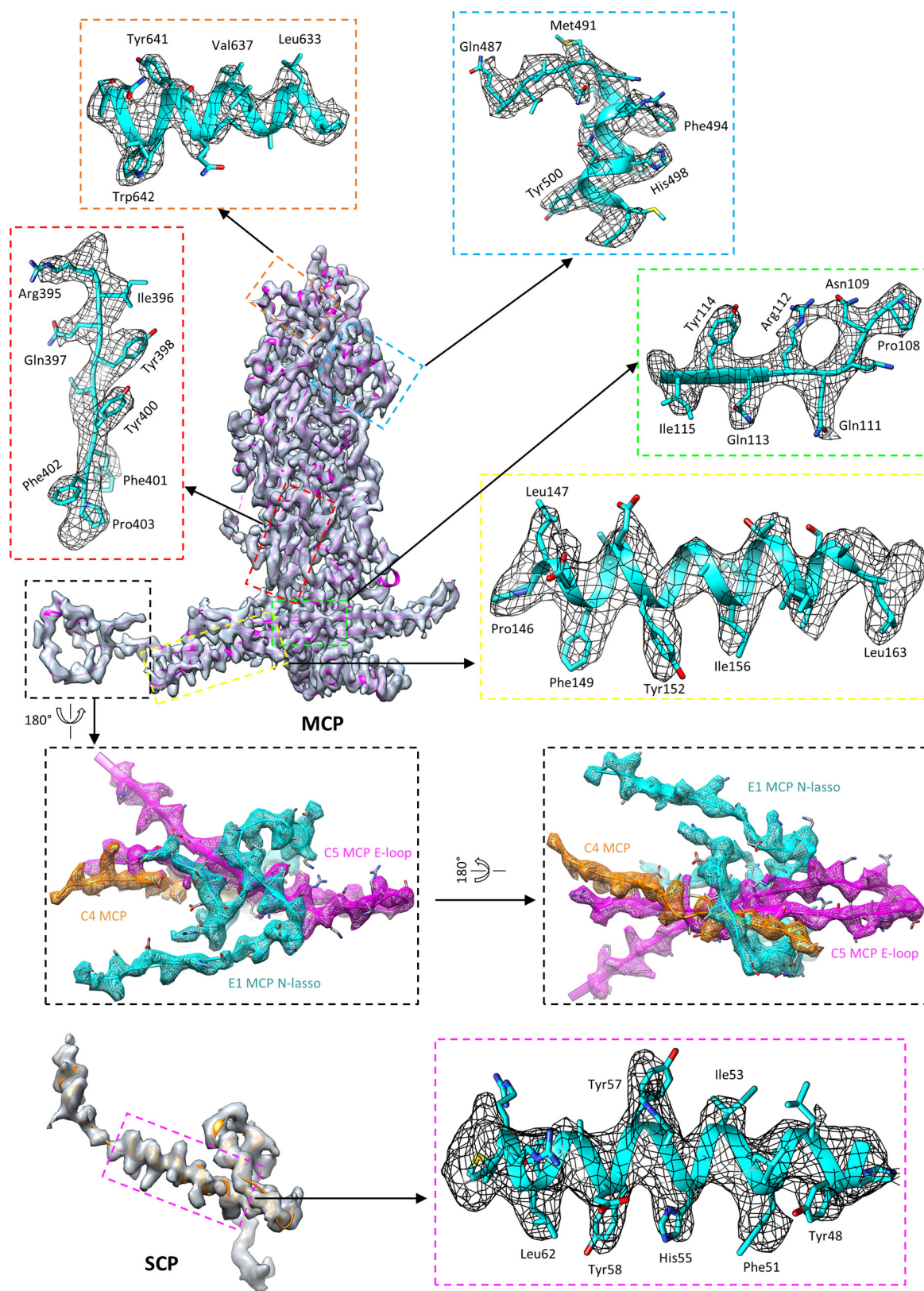


Extended Data Figure 2 | Resolution assessment of the cryo-EM reconstruction and model refinement statistics. **a**, Gold-standard FSC curve of the cryo-EM reconstruction. The average resolution of the final density map is 4.2 Å as determined by the FSC = 0.143 criterion⁴⁴. **b, c**, Local resolution assessment by ResMap⁵². A 640³-voxel sub-volume of the final density map was subjected to ResMap processing. Four slices of

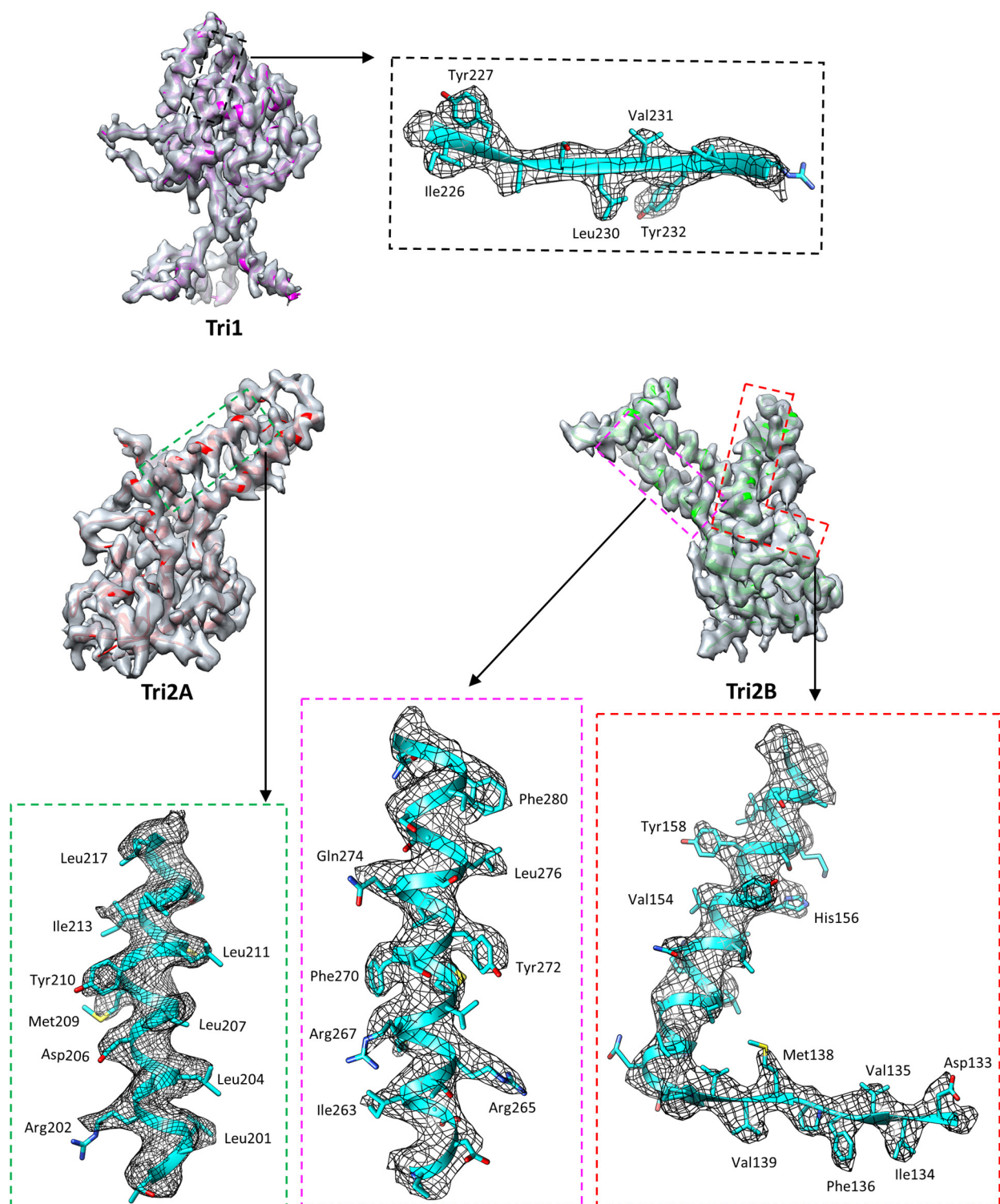
d

Model refinement statistics				
Subunits	Model to map fit (CC around atoms)	Ramachandran plot		
		Favored (%)	Allowed (%)	Outlier (%)
MCP C1	0.7907	93.29	6.64	0.07
MCP C2	0.7631	92.95	6.75	0.30
MCP C3	0.7601	92.87	6.90	0.22
MCP C4	0.7582	92.58	6.97	0.45
MCP C5	0.7769	94.03	5.60	0.37
MCP C6	0.7712	94.22	5.41	0.37
MCP E1	0.7657	92.88	6.97	0.15
MCP E2	0.7690	92.92	6.27	0.81
MCP E3	0.7696	92.80	6.90	0.30
MCP P1	0.7500	91.65	7.87	0.47
MCP P2	0.7569	92.52	6.96	0.52
MCP P3	0.7631	92.43	7.34	0.22
MCP P4	0.7564	92.42	7.21	0.37
MCP P5	0.7483	93.84	5.86	0.30
MCP P6	0.7490	92.29	7.49	0.22
Penton MCP	0.6204	88.30	10.86	0.84
SCP C1	0.7655	97.37	2.63	0.00
SCP C2	0.7747	98.68	1.32	0.00
SCP C3	0.7768	96.05	3.95	0.00
SCP C4	0.7587	98.68	1.32	0.00
SCP C5	0.7761	94.74	5.26	0.00
SCP C6	0.7742	96.05	3.95	0.00
SCP E1	0.7709	93.42	6.58	0.00
SCP E2	0.7834	96.05	3.95	0.00
SCP E3	0.7840	97.37	2.63	0.00
SCP P1	0.7443	94.74	5.26	0.00
SCP P2	0.7356	97.37	2.63	0.00
SCP P3	0.7555	92.11	7.89	0.00
SCP P4	0.7690	97.37	2.63	0.00
SCP P5	0.7472	92.11	7.89	0.00
SCP P6	0.7309	93.42	6.58	0.00
triplex Ta	0.5640	91.10	8.79	0.11
triplex Tb	0.7759	95.56	4.33	0.11
triplex Tc	0.7422	95.09	4.91	0.00
triplex Td	0.7696	95.78	4.00	0.22
triplex Te	0.7607	95.89	4.00	0.11

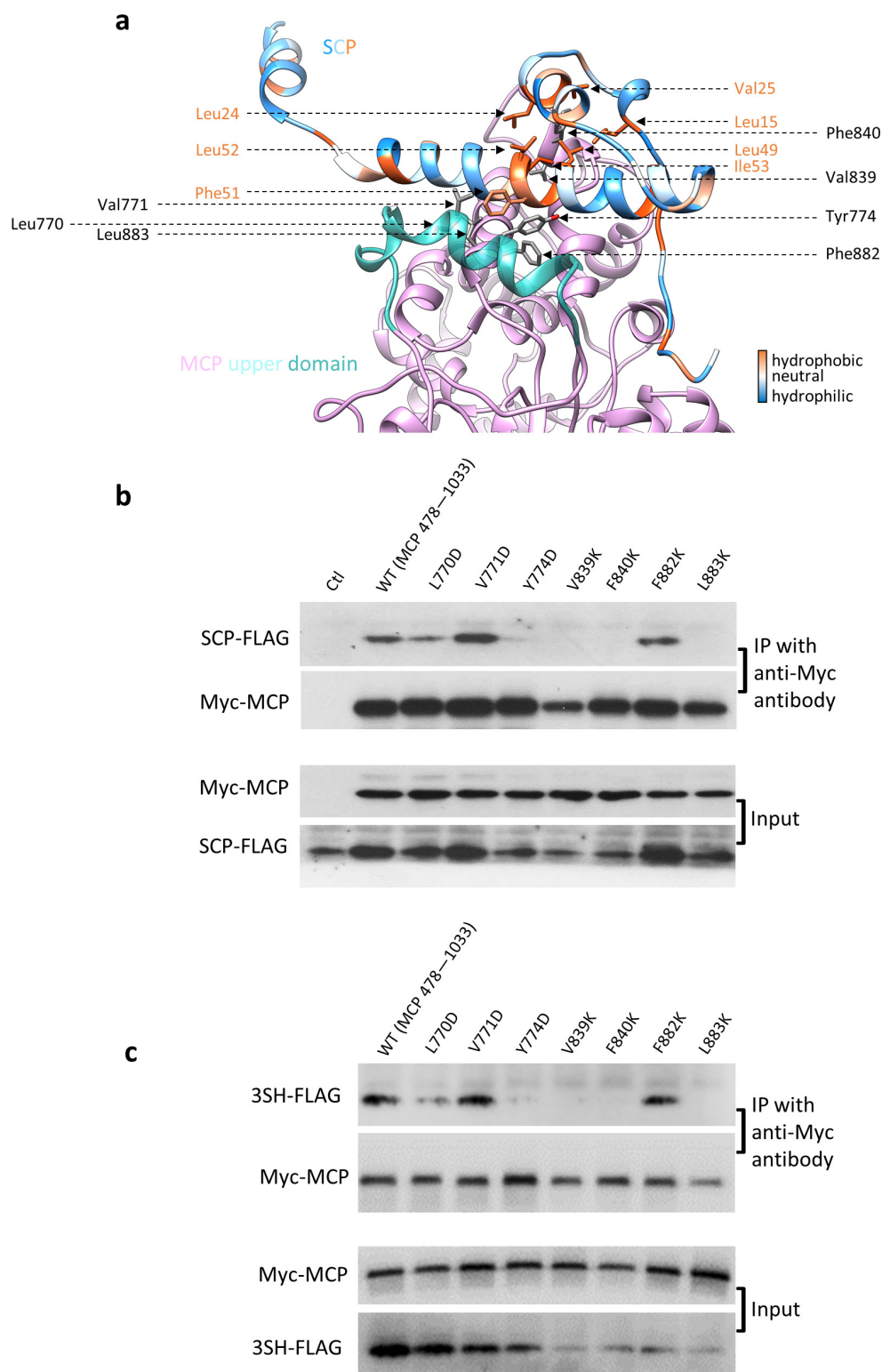
the input volume (**b**) and the local resolution heat map (**c**) are shown. Note that many regions of the density map have better resolution than the FSC-measured average resolution of 4.2 Å. The penton tower region has the lowest resolution because of its flexibility. **d**, Model refinement statistics reported by the Phenix real space refinement program⁴⁷.



Extended Data Figure 3 | Density maps and atomic models of MCP and SCP. Insets correspond to zoomed-in views of boxed regions and illustrate residue features in the density map.



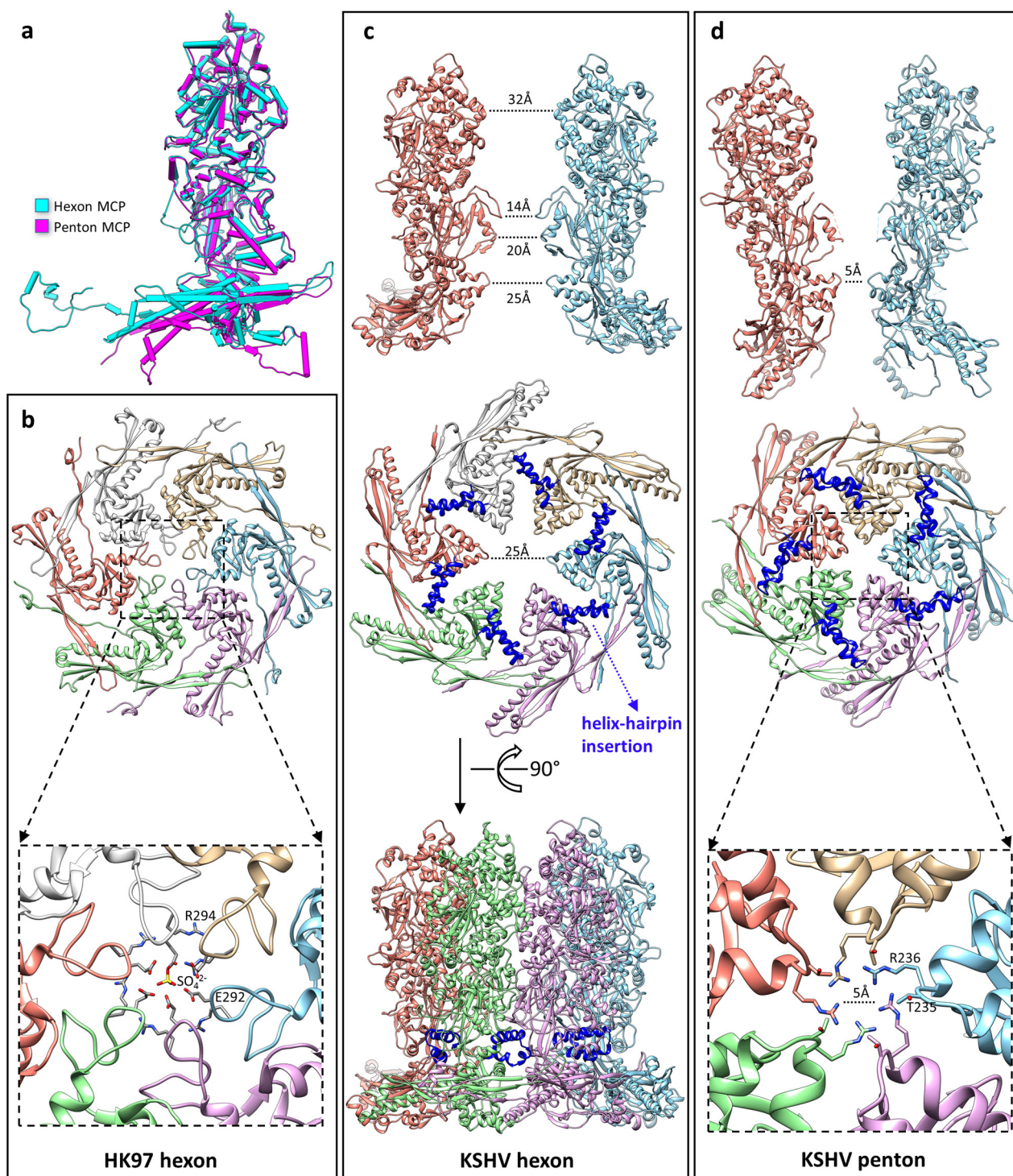
Extended Data Figure 4 | Density maps and atomic models of Tri1, Tri2A and Tri2B. Insets correspond to zoomed-in views of boxed regions and illustrate residue features in the density map.



Extended Data Figure 5 | Structure-guided point mutations of MCPud to identify essential amino acid interactions in MCP–SCP binding.

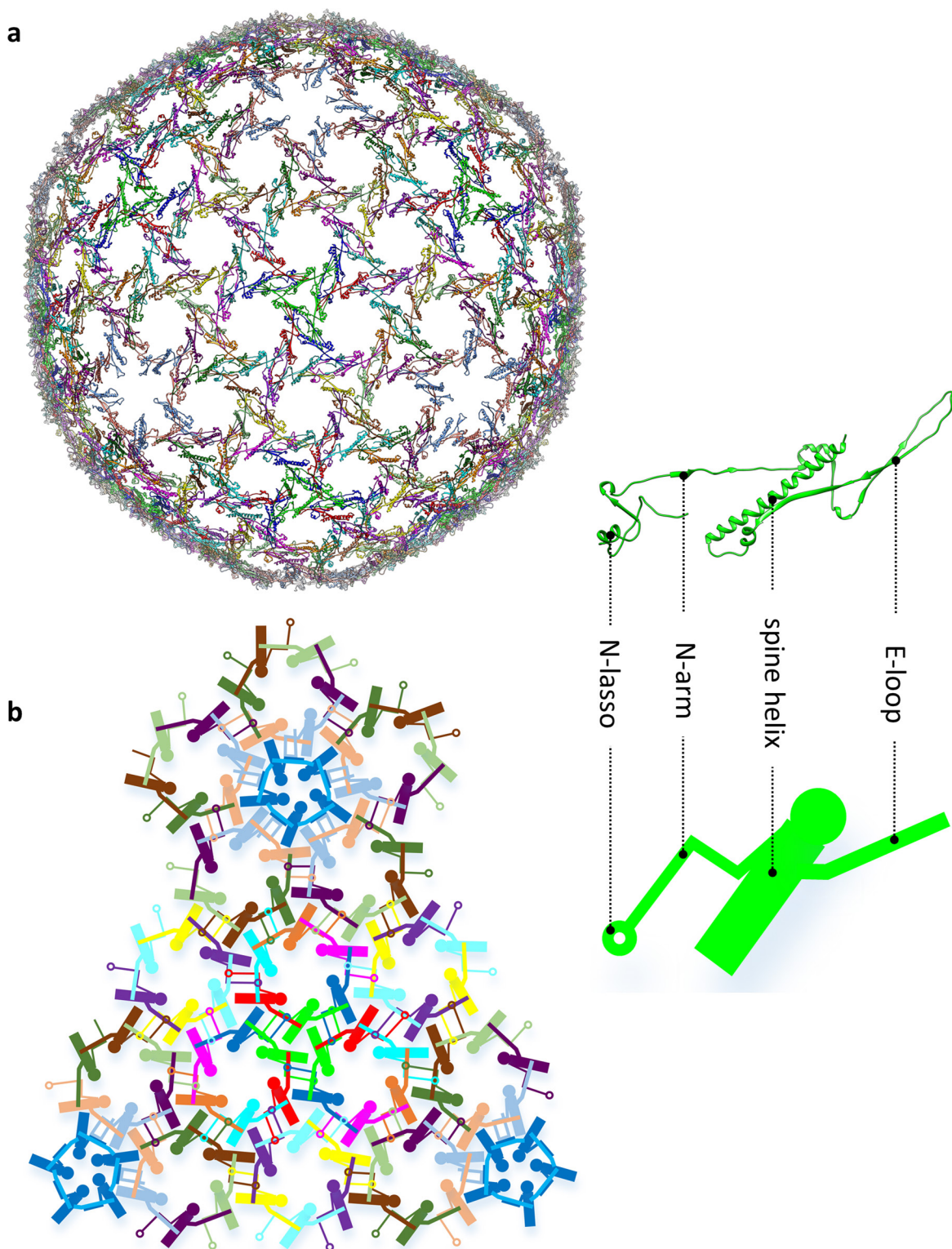
a, Specific amino acid interactions between SCP and MCPud. The SCP model is coloured according to the hydrophobicity of its residues. The MCPud is coloured pink, except for one helix (amino acids 763–778, shown in cyan) that borders the groove for the binding of SCP stem helix. The corresponding region (amino acids 767–781) in the highly homologous HSV-1 MCPud is a loop structure, which thus forms a relatively flat surface for SCP binding. **b**, **c**, Demonstration of specific interaction between SCP and MCPud (**b**) or between the SCP-mimicking polypeptide 3SH–Flag and MCPud (**c**); 293T cells were co-transfected with expression plasmids of MCPud (amino acids 478–1033, wild type

or mutants) and SCP–Flag (**b**) or of MCPud and 3SH–Flag (**c**). Two days later, cell lysates were subjected to a co-immunoprecipitation assay using mouse anti-Myc antibody, and further analysed by western blotting with rabbit anti-Flag and anti-Myc antibodies. As expected, expressed wild-type SCP (**b**) and 3SH–Flag (**c**) both bound to expressed wild-type MCPud. Conversely, four out of the seven MCPud point mutations that substitute a hydrophobic residue with a hydrophilic residue disrupted these interactions for both wild-type SCP (**b**) and 3SH–Flag (**c**). These results suggest that SCP-mimicking polypeptides interact with MCPud in a similar way to wild-type SCP. Experiments were repeated independently twice with similar results.



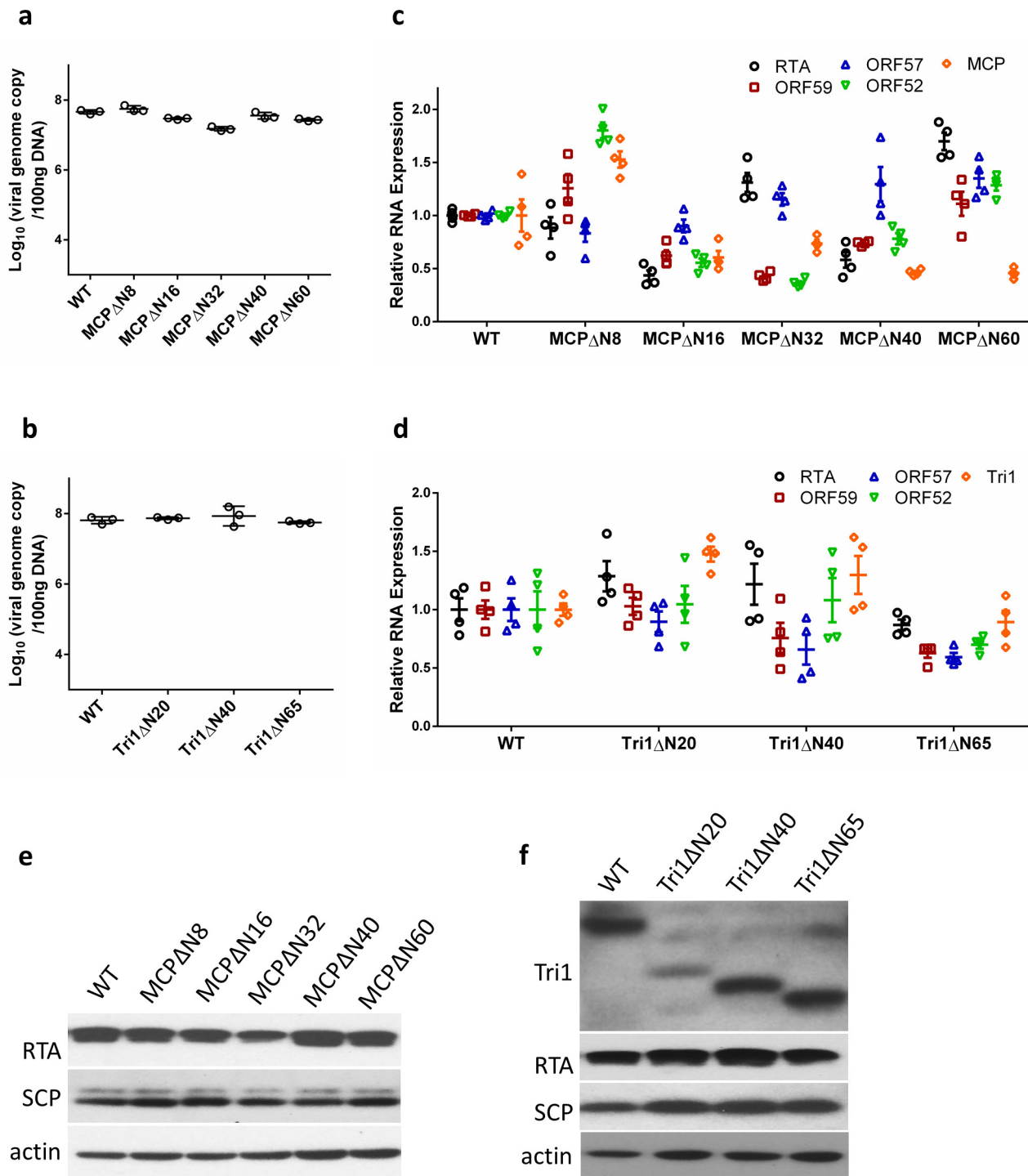
Extended Data Figure 6 | Structural differences between hexon and penton. **a**, Superimposed models of a hexon MCP and a penton MCP, which shows the hinged tilting of the floor region of a penton MCP towards the centre of the capsid. **b–d**, Comparison of channel constrictions between a bacteriophage HK97 hexon (**b**), a KSHV hexon (**c**) and a KSHV penton (**d**). The hexon channel in HK97 is tightly constricted by a loop in the A-domain (Fig. 2k) of the Johnson fold (**b**, top). In the crystal structure of HK97, this channel is completely blocked by a sulfate ion⁵³ (**b**, bottom). The hexon channel in KSHV is not constricted by the HK97-like Johnson fold. The diameter of the channel at this position is 25 Å (**c**, middle). Note the large gap between adjacent Johnson folds, and

also the absence of a long loop corresponding to the channel-constricting loop in HK97. Instead, the KSHV hexon channel is most constricted at the channel domain (**c**, top). Side view of the hexon (**c**, bottom) shows that the helix-hairpin domain insertion (blue) seals a hole at the root of the capsomer tower. The penton channel in KSHV is constricted by the Johnson-fold domain (**d**, middle). The diameter of the penton channel at this position is 5 Å (**d**, bottom). Owing to the hinged tilting of the floor region of penton MCP towards the centre of the capsid (**a**), adjacent Johnson-fold domains move closer to one another and constrict the penton channel (**d**, top).



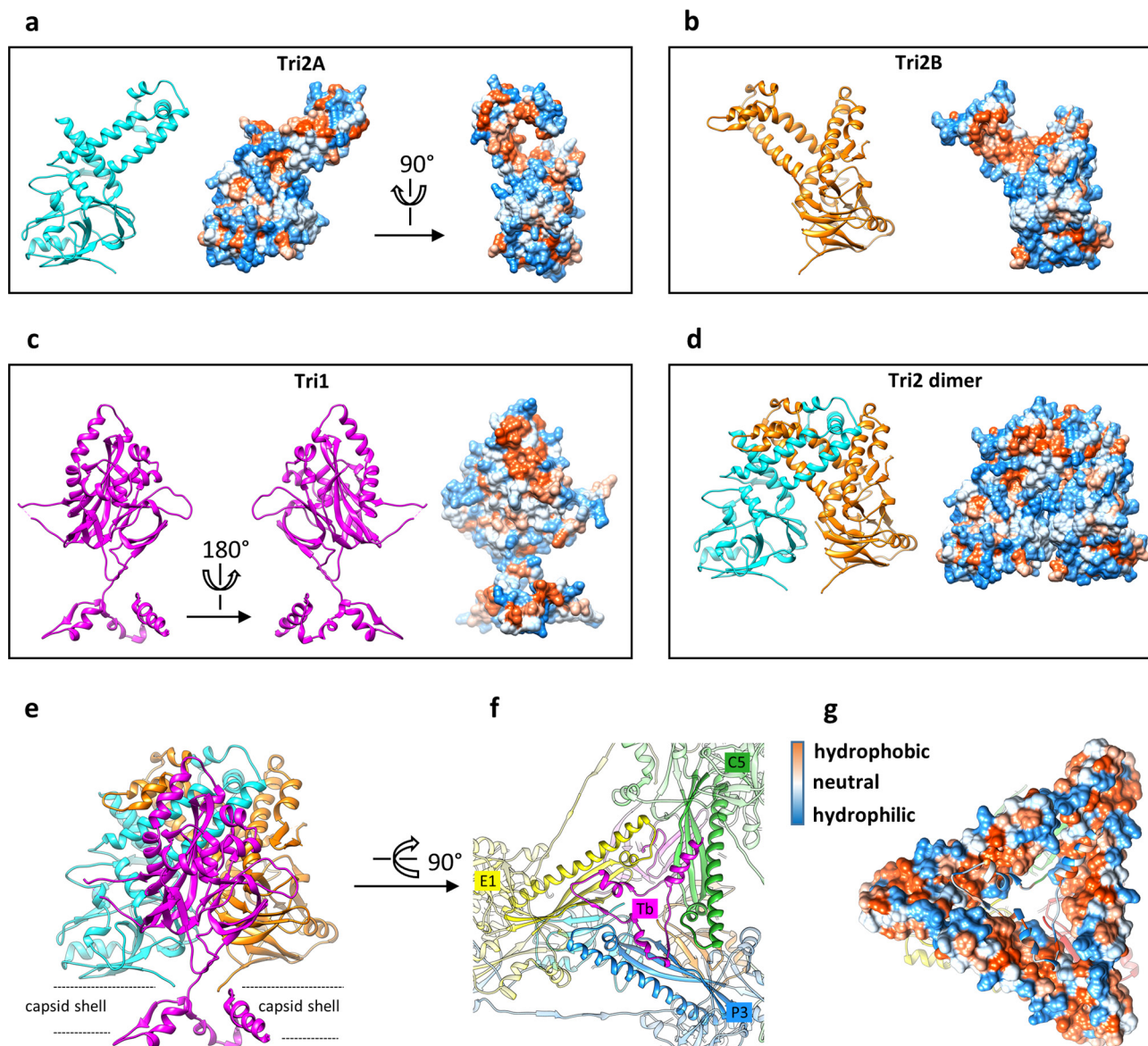
Extended Data Figure 7 | The MCP network in KSHV capsid. **a**, The network forged by MCP N-lassoes in the KSHV capsid floor, illustrated with atomic models. A short continuous segment of MCP N-terminal region (amino acids 1–186) including the N-lasso, N-arm, E-loop and

spine helix is shown in each MCP model. **b**, A schematic representation of part of the network. An analogy can be drawn between the ‘dancer’ in the schematic representation and the MCP atomic model as shown in the inset.



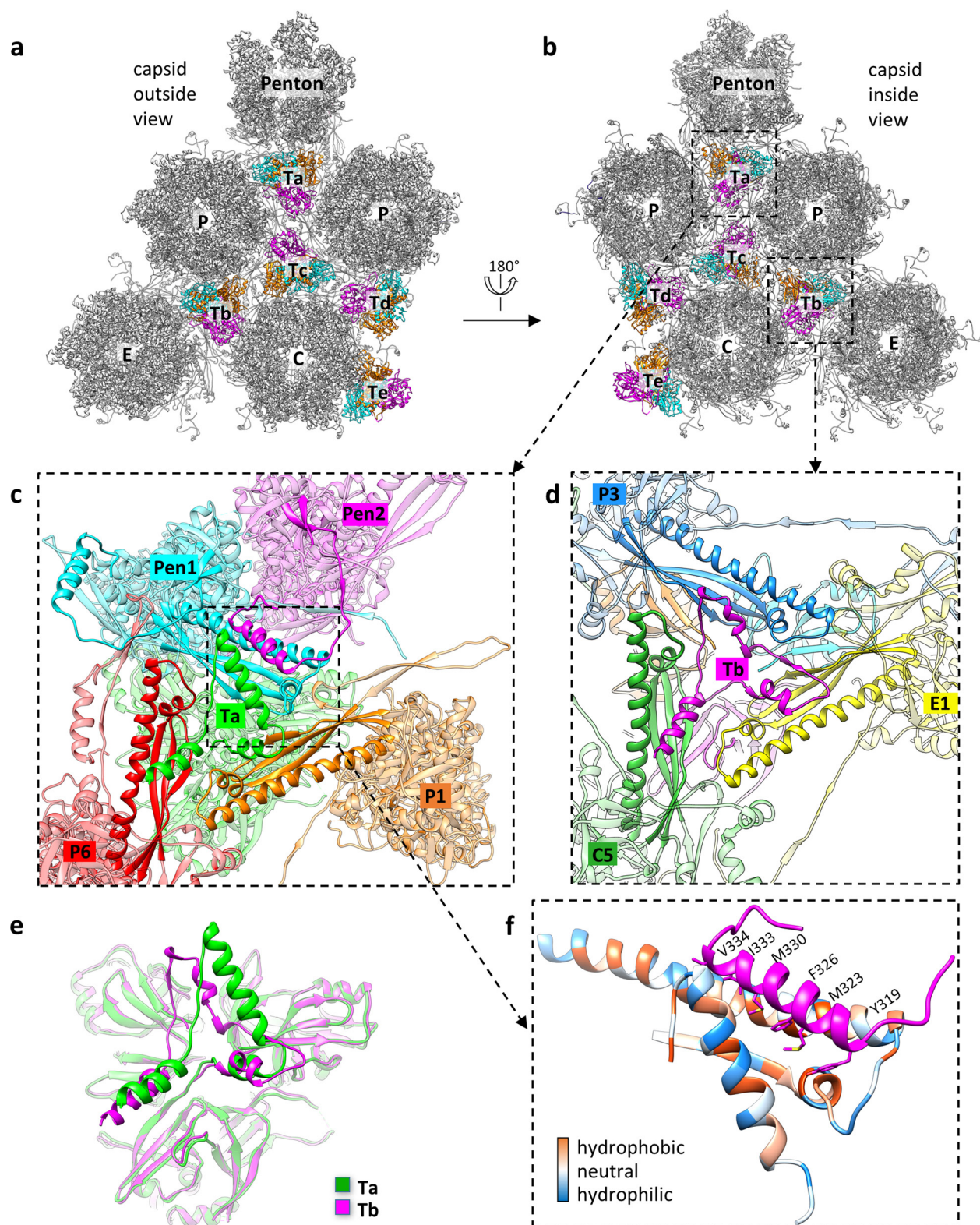
Extended Data Figure 8 | Truncation mutagenesis of the MCP N-lasso or Tri1 N-anchor does not notably affect KSHV DNA replication or gene expression. **a, b**, Viral genome copy number in cells replicating the wild type (WT), MCP-truncated (**a**) or Tri1-truncated (**b**) KSHV. Design of MCP truncations or Tri1 truncations is shown in Figs 3i and 4j, respectively. KSHV lytic replication was induced in cells harbouring the wild-type or the mutated KSHV genome. Total DNA was extracted from cells, and viral genome copy number was determined by real-time PCR. Data are mean \pm s.e.m. ($n = 3$ biologically independent samples). **c, d**, Viral RNA expression in cells replicating the wild-type, MCP-truncated (**c**) or Tri1-truncated (**d**) KSHV. Total RNA was

extracted from cells induced for KSHV lytic replication. Viral RNA transcripts were quantified by real-time PCR with reverse transcription and presented as fold changes over RNA level of wild-type virus. Data are mean \pm s.e.m. ($n = 3$ or 4 biologically independent samples). **e, f**, Expression of viral and cellular proteins in cells replicating the wild-type, MCP-truncated (**e**) or Tri1-truncated (**f**) KSHV. Correct sizes of truncated Tri1 were verified by western blotting with an anti-Tri1 antibody as shown in **f**. Verification of truncated MCP was not carried out owing to the lack of anti-MCP antibody. Experiments were repeated independently twice with similar results.



Extended Data Figure 9 | Hydrophobic interactions in the formation of triplexes and in the anchoring of triplexes to the capsid floor. a–d, Hydrophobic interactions have a major role in the formation of triplex heterotrimers. Surface representations of Tri2A (a), Tri2B (b) and Tri1 (c) monomers or the Tri2 dimer (d) were calculated and coloured according to hydrophobicity. Red, hydrophobic; white, neutral; blue,

hydrophilic. Large patches of hydrophobic residues at the interface of the Tri2A and Tri2B embracing arm domains hold the Tri2 dimer together, and contribute to interactions with the Tri1 third-wheel domain to form the heterotrimer. e–g, Triplexes are anchored to the capsid floor by the tripod-shaped Tri1 N-anchor (e, f) via hydrophobic interactions (g).



Extended Data Figure 10 | Structural polymorphism in the Tri1 N-anchor. **a, b**, Distribution of triplexes in the MCP network viewed from outside (**a**) or inside (**b**) the capsid. **c, d**, Zoomed-in views of triplex Ta (**c**) or Tb (**d**) from inside the capsid. **e**, Superimposed models of triplexes Ta and Tb reveal structural differences in their Tri1 N-anchor domains.

f, The refolded Tri1 N-anchor in triplex Ta contributes to penton stabilization. The refolded helix in Ta Tri1 forms a hydrophobic cleft with the spine helix of a penton MCP, in which the refolded dimerization domain of an adjacent penton MCP (magenta) binds with a series of hydrophobic residues.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	KSHV capsid (EMD-7047) (PDB 6B43)
Data collection and processing	
Magnification	14,000
Voltage (kV)	300
Electron exposure (e ⁻ /Å ²)	25
Defocus range (μm)	1-3
Pixel size (Å)	1.03
Symmetry imposed	Icosahedral
Initial particle images (no.)	44,343
Final particle images (no.)	29,100
Map resolution (Å)	0.143
FSC threshold	
Map resolution range (Å)	4.2
Refinement	
Initial model used (PDB code)	N/A
Model resolution (Å)	N/A
FSC threshold	
Model resolution range (Å)	N/A
Map sharpening <i>B</i> factor (Å ²)	200
Model composition	
Non-hydrogen atoms	214,334
Protein residues	27,223
Ligands	0
<i>B</i> factors (Å ²)	N/A
Protein	
Ligand	
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.119
Validation	
MolProbity score	1.98
Clashscore	7.41
Poor rotamers (%)	0.08
Ramachandran plot	
Favored (%)	89.26
Allowed (%)	10.59
Disallowed (%)	0.15

Atomic structure of the eukaryotic intramembrane RAS methyltransferase ICMT

Melinda M. Diver^{1,2†}, Leanne Pedi¹, Akiko Koide^{3,4}, Shohei Koide^{3,5} & Stephen B. Long¹

The maturation of RAS GTPases and approximately 200 other cellular CAAX proteins involves three enzymatic steps: addition of a farnesyl or geranylgeranyl prenyl lipid to the cysteine (C) in the C-terminal CAAX motif, proteolytic cleavage of the AAX residues and methylation of the exposed prenylcysteine residue at its terminal carboxylate¹. This final step is catalysed by isoprenylcysteine carboxyl methyltransferase (ICMT), a eukaryote-specific integral membrane enzyme that resides in the endoplasmic reticulum². ICMT is the only cellular enzyme that is known to methylate prenylcysteine substrates; methylation is important for the biological functions of these substrates, such as the membrane localization and subsequent activity of RAS¹, prelamin A³ and RAB⁴. Inhibition of ICMT has potential for combating progeria³ and cancer^{5–8}. Here we present an X-ray structure of ICMT, in complex with its cofactor, an ordered lipid molecule and a monobody inhibitor, at 2.3 Å resolution. The active site spans cytosolic and membrane-exposed regions, indicating distinct entry routes for the cytosolic methyl donor, S-adenosyl-L-methionine, and for prenylcysteine substrates, which are associated with the endoplasmic reticulum membrane. The structure suggests how ICMT overcomes the topographical challenge and unfavourable energetics of bringing two reactants that have different cellular localizations together in a membrane environment—a relatively uncharacterized but defining feature of many integral membrane enzymes.

ICMT from the beetle *Tribolium castaneum* exhibited superior biochemical stability in detergent-containing solutions in comparison to other orthologues and was used for biochemical characterization and structure determination (see Methods). Human and beetle ICMT share the same predicted topology⁹ and have 58% amino acid sequence identity within the region thought to contain the active site¹⁰ (amino acids 90–281; Extended Data Fig. 1). Beetle ICMT demonstrated robust methylation of prenylcysteine substrates both in cellular membranes and in the purified form, and exhibited kinetic parameters similar to those of human ICMT^{11,12} (Extended Data Fig. 2). We engineered a synthetic ICMT-binding protein called a ‘monobody’, based on a randomized fibronectin protein domain, for use as a crystallization chaperone¹³. The monobody is an inhibitor of ICMT ($IC_{50} \approx 1 \mu M$), and exhibits specificity for the beetle orthologue (Extended Data Fig. 2g, h). Crystals of purified ICMT–monobody complex were obtained in the lipidic-cubic phase¹⁴ in the presence of the S-adenosyl-L-homocysteine (AdoHcy) cofactor and the prenylcysteine substrate N-acetyl-S-geranylgeranyl-L-cysteine (AGGC) (Extended Data Fig. 2a). Experimental phases yielded high-quality electron-density maps that enabled the placement of all amino acids of ICMT and the monobody (Extended Data Fig. 3a). The refined atomic coordinates have good stereochemistry and an R_{free} value of 24.6% (Extended Data Table 1).

ICMT contains eight transmembrane α -helices (M1–M8) and would reside almost entirely within the endoplasmic reticulum membrane in a

cellular context (Fig. 1a, b). The largest cytosolic region of the enzyme, which protrudes approximately 12 Å away from the membrane and encompasses the binding site for AdoHcy, is formed by an extension of M8 together with a structurally ordered connection between M6 and M7 (the M6–M7 connector) plus a short ‘cap’ helix near the C terminus. The M6–M7 connector does not fully reach the luminal side, but is stabilized within the transmembrane region by interactions with the M5–M6 connector, which lies beneath it (Fig. 1b and Extended Data Fig. 4a). Additionally, the M5 helix would not span the bilayer. Its N terminus, capped by a hydrogen bond with Ser128, is positioned within the membrane region, about 10 Å from the cytosolic side (Fig. 1 and Extended Data Fig. 5b). Unusually, the M4 and M5 helices are connected by a 25 Å-long extended segment (Pro115–Pro129) that traverses diagonally in the membrane-spanning region (Fig. 1c and Extended Data Fig. 5a). The M1–M3 region, which is unique to ICMT enzymes from animals (Extended Data Fig. 1), makes extensive contacts with M4 and the M4–M5 connector. Helices M1, M2 and M3 associate with one another and are stabilized, in part, by GXXXG-like helical packing motifs between M1 and M3¹⁵ (Extended Data Fig. 6a). The non-covalent association is strong enough that M1 and M2 remain associated with the core of the enzyme following proteolytic cleavage of the M2–M3 loop (Extended Data Fig. 6). Congruently, genetic deletion of M1 and M2 renders human ICMT inactive¹².

The structure provides context for a large body of functional data on ICMT^{9–12,16}. Residues that disrupt catalytic activity when mutated define the active site on the structure (Fig. 2a), the location of which had remained elusive owing to the broad distribution of these residues in the primary structure (Extended Data Fig. 1). In the cell, the active site would be located mostly within the cytosolic leaflet of the membrane, and is contained between M4 and the C terminus in the primary structure (Fig. 2a). AdoHcy is encapsulated within a pocket that secludes it from both the aqueous environment of the cytosol and the lipid membrane, and is formed by the M6–M7 connector, the cytosolic extension of M8 and the cap helix (Fig. 2b, c). The release of AdoHcy and the subsequent binding of S-adenosyl-L-methionine (AdoMet) could occur via hinged displacement of the M6–M7 connector towards the cytosol, in which there is a reservoir of micromolar concentrations of AdoMet; Gly181 and Ser193 are potential hinge points for this mechanism (Fig. 2c). The perfectly conserved residues Phe184, Tyr204 and Glu250 appear to be particularly important for positioning AdoHcy through direct contacts (Fig. 2c), and completely abolish activity when mutated to alanine^{12,16} (Extended Data Fig. 1). The active site could accommodate the methyl group of AdoMet in a slender ‘tunnel’ (Fig. 2b and Extended Data Fig. 4c), suggesting that the observed conformation of AdoHcy is analogous to the conformation when the methyl donor is bound.

ICMT accommodates a breadth of protein substrates. This includes all prenylated (farnesylated and geranylgeranylated) and proteolytically

¹Structural Biology Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. ²Graduate Program in Biochemistry and Structural Biology, Cell and Developmental Biology, and Molecular Biology, Weill Cornell Medicine Graduate School of Medical Sciences, 1300 York Avenue, New York, New York 10065, USA. ³Perlmutter Cancer Center, New York University Langone Medical Center, 430 East 29th Street, New York, New York 10016, USA. ⁴Department of Medicine, New York University Langone Medical Center, 430 East 29th Street, New York, New York 10016, USA. ⁵Department of Biochemistry and Molecular Pharmacology, New York University Langone Medical Center, 430 East 29th Street, New York, New York 10016, USA. [†]Present address: Department of Physiology, University of California, San Francisco, San Francisco, California 94158, USA.

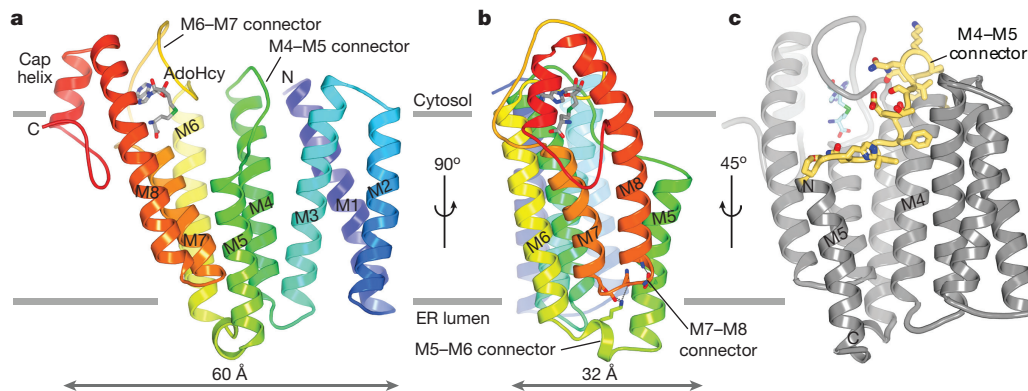


Figure 1 | Architecture of ICMT. **a**, Ribbon representation with secondary structure elements coloured blue-to-red from N terminus to C terminus. Horizontal lines indicate approximate boundaries of the endoplasmic reticulum (ER) membrane. **b**, Orthogonal view, showing interactions

between the M5–M6 and M7–M8 connectors (sticks with hydrogen bonds drawn as dashed lines). **c**, View highlighting the location of the M4–M5 connector (yellow with drawn side chains).

processed CAAX proteins, as well certain RAB GTPases that have two geranylgeranyl groups attached to the cysteine residues of their C-terminal CXC motif¹. The structure suggests that the prenyl moiety of the substrate binds in a deep cavity within the active site that extends from the cofactor pocket into the transmembrane region (Figs 2b and 3a). This cavity is approximately 22 Å long and 6 Å wide, and is formed by regions of the M4, M5, M7 and M8 helices. Approximately two thirds of the cavity would be exposed to the cytosolic leaflet of the lipid bilayer via a lateral crevice between helices M5 and M8 (Figs 2b and 3b). A tube of electron density consistent with a lipid is present in the cavity (Fig. 3a). Although the 20-carbon geranylgeranyl group of the AGGC substrate that was present during crystallization could account for the density, we cannot be sure of its identity, and therefore modelled a monoolein lipid into the density (Fig. 3a and Extended Data Fig. 3b). The cavity would also accommodate the 15-carbon farnesyl group. Similar to the binding site for the prenyl group in water-soluble enzymes such as farnesyltransferase¹⁷, the proposed prenyl-binding cavity is lined primarily by aromatic amino acids (Tyr95, Met99, Phe102, Val124, Asn126, Tyr131, Trp215, Trp218, Tyr235, Phe242 and Phe243) that markedly reduce enzyme activity when mutated¹² (Extended Data Fig. 1). For doubly geranylgeranylated CXC proteins, the second prenyl group could be accommodated in the crevice between M5 and M8 and/or in the adjacent membrane while the C-terminal geranylgeranyl group occupies the active site.

The structure of a methyltransferase from the prokaryotic organism *Methanosarcina acetivorans*, which we refer to as *MaMTase*, highlights

the diversity within the integral membrane methyltransferase family¹⁶ (Extended Data Fig. 7). Despite sharing only 14% sequence identity with ICMT, *MaMTase* was previously referred to as *Ma-ICMT*, but this nomenclature was misleading because its biological substrates are unknown, it cannot methylate prenylated-peptide substrates¹⁶ and there are no known prenylated proteins in prokaryotes. Nevertheless, the binding sites for AdoHcy are analogous in ICMT and *MaMTase* (Extended Data Fig. 7) and the cofactor-binding domain of ICMT, which spans from M6 through the cap helix, shares a recognizable fold not only with regions of *MaMTase*, but also with regions of a prokaryotic integral membrane sterol reductase that utilizes a nicotinamide adenine dinucleotide phosphate (NADP⁺) cofactor^{16,18}; this suggests that this domain represents a structural motif for soluble cofactor binding to integral membrane enzymes (Extended Data Fig. 8). Other regions of the active site of ICMT show minimal similarity to *MaMTase*; the substrate-binding sites differ in size, amino acid composition and membrane exposure (Fig. 3b, Extended Data Figs 7a and 8).

The monobody inhibitor binds ICMT adjacent to the active site and interacts with portions of M5, M8 and the M6–M7 loop (Fig. 3c). The ‘FG loop’ of the monobody, which is diversified in the combinatorial library¹³, dips into the membrane region and presents a tryptophan residue (Trp80) that occupies part of the crevice between M5 and M8 (Fig. 3d). It also contacts a portion of the modelled lipid. Although monobodies typically recognize native surfaces of their cognate proteins¹⁹, to evaluate whether the structure of ICMT was affected

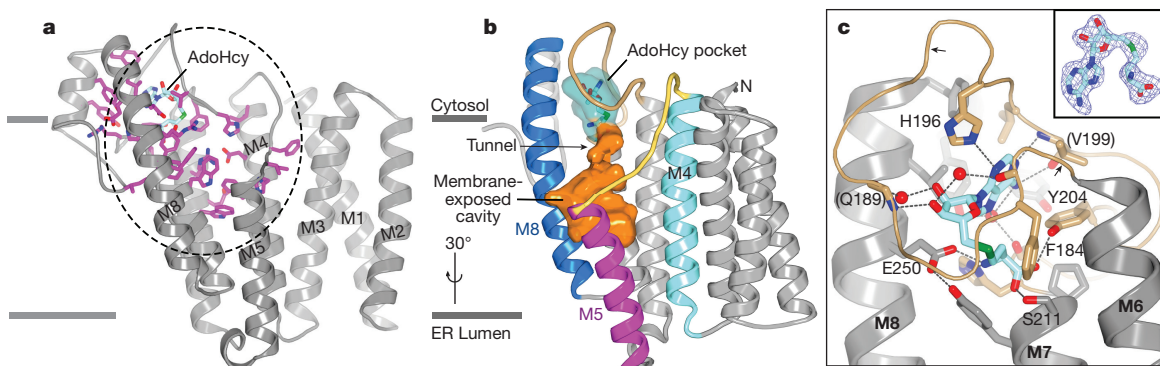


Figure 2 | The active site. **a**, Three-dimensional clustering of key residues identifies the active site. Residues that diminish specific activity by more than 95% when mutated¹² (Extended Data Fig. 1) are drawn in magenta on a ribbon representation. The high concentration of such residues within the dashed region demarcates the active site. **b**, The active site, depicted as a molecular surface within the ribbon representation. The AdoHcy pocket is cyan; the membrane-exposed cavity is orange. **c**, Interactions with

AdoHcy (sticks with cyan carbon atoms). Dashed lines indicate hydrogen bonds. Water molecules are shown as red spheres. Nitrogen, blue; oxygen, red; and sulfur, green. The M6–M7 connector is tan. Arrows mark the locations of Gly181 and Ser193; parentheses indicate hydrogen bonds with backbone atoms. Inset shows electron density for AdoHcy (simulated annealing omit $F_o - F_c$ map, contoured at 3σ).

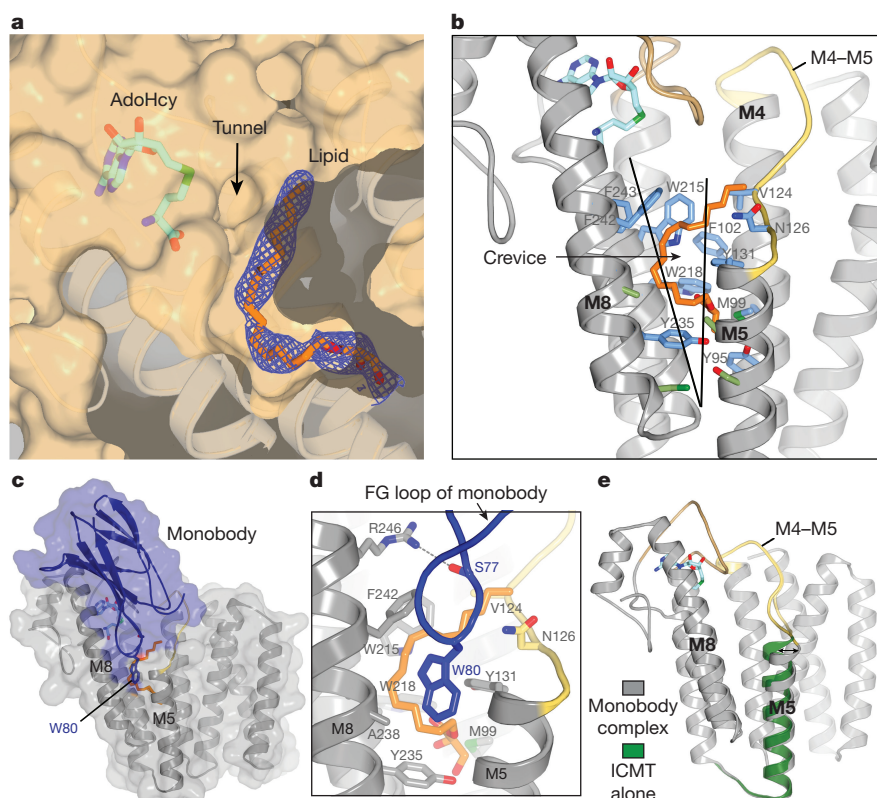


Figure 3 | Lipid-binding cavity and monobody complex. **a**, The active-site cavity containing a lipid molecule. Electron density (blue mesh, $2F_o - F_c$ map contoured at 1σ) for a monoolein lipid (orange sticks) is shown on a cutaway view of the molecular surface (tan). The tunnel between AdoHcy (sticks) and the lipid-binding cavity is indicated. Helices below the surface are depicted as ribbons. **b**, The crevice and amino acids comprising the lipid-binding cavity. A portion of ICMT is depicted as ribbons with the M4–M5 connector in yellow and the M6–M7 connector in tan. The crevice between M5 and M8 is indicated by a wedge. Amino acids within van der Waals distances of the lipid are coloured blue; those that line the crevice

are green. **c**, Monobody complex. A semi-transparent representation of the molecular surface of the complex (ICMT, grey; monobody, blue) is shown surrounding ribbon representations. **d**, Close view of the FG loop of the monobody (blue) interacting with ICMT (predominately grey). Trp80 and Ser77 of the monobody are shown as sticks. A hydrogen bond between Ser77 and Arg246 of ICMT is shown as a dashed line. The monoolein molecule (orange) and ICMT residues surrounding it are shown as sticks. **e**, Comparison of structures with and without the monobody (coloured as indicated). An arrow highlights the tilting of M5; the overall structures are otherwise indistinguishable.

by the monobody, we determined a 4.0 Å resolution X-ray structure of ICMT without the monobody (Methods). The only discernable difference is a slight (around 5°) tilting of M5, which could also be due to crystallization in detergent rather than lipidic cubic phase and/or a degree of inherent flexibility (Fig. 3e and Extended Data Fig. 3e; the overall r.m.s.d. for Cα atoms is 0.5 Å). The lengthy lipid-binding cavity and the crevice between M5 and M8 leading to it, hallmark features of the active site, are present without the monobody. The location of the monobody suggests that it would prevent prenylated substrates from reaching the active site and/or block product release. Hydrogen bonding between Ser77 of the monobody and Arg246 of ICMT may also contribute to its inhibitory function (Fig. 3d), as we predict that Arg246 coordinates the C-terminal carboxylate of the prenylcysteine. Although specific contacts between ICMT and cognate protein substrates are expected to be confined to their prenylcysteine moieties^{20–22}, the positioning of the monobody adjacent to the active site and on the cytosolic side above the crevice between M5 and M8 may demarcate the approximate location of a protein substrate as its C-terminal prenylcysteine undergoes methylation (Figs 3c and 4a). Hydrophobic molecules that bind in the crevice would be expected to inhibit the enzyme; this may be the mode of action of some existing ICMT inhibitors.

The protein substrates of ICMT are amphipathic; their prenyl group or groups partition into the endoplasmic reticulum membrane¹, and their C-terminal carboxylate and variable protein portions are hydrophilic and exposed to the cytosol. The crevice between M5 and M8, which would be accessible to the hydrophobic core of the membrane

and leads directly to the active site, provides a plausible route for prenyl entry by lateral diffusion (Fig. 4a). The positions of the N-terminal end of M5 and the M4–M5 connector create a hydrophilic depression within the membrane-embedded region of ICMT, and might induce a concomitant depression in the proximal lipids of the bilayer (Extended Data Figs 3c and 9) that would accommodate the upstream peptide, as has been modelled for KRAS4B (Fig. 4a). Suggestive of an important role, the M4–M5 connector is highly conserved and greatly diminishes catalytic activity when mutated¹² (Fig. 2a and Extended Data Fig. 1). Hydrophobic residues (Phe123, Val124 and Leu125) presumably anchor it in the membrane, and hydrophilic residues (Asn126, His127 and Ser128) give it amphipathic character (Extended Data Fig. 5a). The enzyme appears to achieve substrate specificity by the distinct position, shape, proportion and amphiphilicity of its active site.

We constructed a model of a transition state based on the structure of ICMT (Fig. 4b). The negatively charged carboxylate of the prenylcysteine substrate, the nucleophile in the direct-transfer mechanism, is coordinated and positioned for catalysis by two arginine residues, Arg173 (on M6) and Arg246 (on M8), which also provide specificity for the carboxylate. In the X-ray structure, Arg173 and Arg246 are stabilized by hydrogen-bonding networks and two water molecules occupy the proposed location of the carboxylate (Extended Data Fig. 4b). Similar to other methyltransferases that use AdoMet as the methyl donor²³, and supported by mutagenesis data¹² (Extended Data Fig. 1), we predict that the methyl group of AdoMet makes three unconventional hydrogen bonds involving carbon as the hydrogen-bond donor that help stabilize the transition state (Fig. 4b).

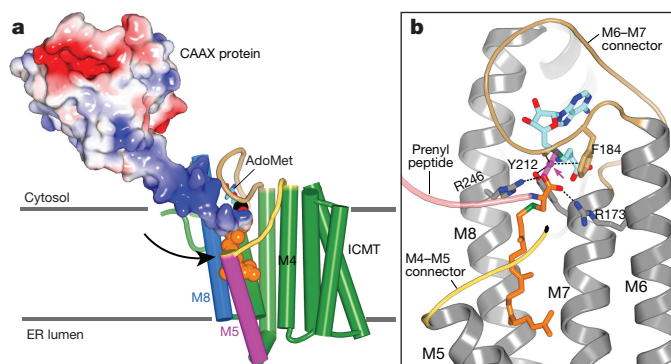


Figure 4 | Substrate access and transition-state models. **a**, Substrate access in a hypothetical ternary complex. ICMT is depicted with cylindrical helices; AdoMet is modelled from AdoHcy. The binding of an ICMT substrate, in this case partially processed KRAS4B (coloured molecular surface; RCSB Protein Data Bank (PDB) ID: 5TAR) with attached geranylgeranyl group (orange spheres), is shown. An arrow denotes a plausible path for membrane-partitioned substrates to reach the active site through the crevice between M5 and M8. The hypervariable region of the substrate upstream of the CAAX motif is dark blue, indicating its basic character in KRAS4B. **b**, Transition-state model. Translucent magenta sticks depict the scissile and nascent bonds, with the methyl group of AdoMet (cyan carbons) shown as a magenta sphere (indicated by the arrow). The C-terminal carboxylate of the prenylcysteine substrate (orange carbons) forms hydrogen bonds (dashes) with Arg173 and Arg246 that orient it for inline nucleophilic attack. The transition state is further stabilized by interactions (dashes) with the methyl group being transferred: a CH–O hydrogen bond with the side chain hydroxyl of Tyr212, a CH–O hydrogen bond with the backbone carbonyl oxygen of Asn185, and a CH– π interaction with the aromatic face of Phe184. All of the amino acid side chains proposed to be involved in transition state stabilization are perfectly conserved among ICMT enzymes. In the figure, the peptide portion of a hypothetical substrate is drawn as a pink ribbon, a portion of the M4–M5 connector is coloured yellow, and the enzyme is shown as grey ribbons; M1–M4 are removed for clarity.

A recent study indicates that human ICMT has a synthetic lethal interaction with oncogenic RAS⁸, suggesting that ICMT inhibitors could be effective against RAS-driven cancers. ICMT inhibitors may also be useful for treating progeria, the pathology of which is attributed to the accumulation of a prenylated and methylated form of prelamin A at the nuclear envelope^{3,24}. This structure of ICMT will contribute toward efforts to develop inhibitors. More generally, our study describes the atomic structure of an intramembrane enzyme, a class of proteins for which a relatively small number of structures have been determined. Many enzymes that are embedded in biological membranes facilitate the access of reactants that have drastically different physicochemical properties to a common active site and maintain them in close proximity for catalysis. The structure provides insight into how one enzyme accomplishes these complex tasks while maintaining specificity for its diverse substrates.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 September; accepted 7 December 2017.

Published online 17 January 2018.

- Wang, M. & Casey, P. J. Protein prenylation: unique fats make their mark on biology. *Nat. Rev. Mol. Cell Biol.* **17**, 110–122 (2016).
- Dai, Q. *et al.* Mammalian prenylcysteine carboxyl methyltransferase is in the endoplasmic reticulum. *J. Biol. Chem.* **273**, 15030–15034 (1998).
- Ibrahim, M. X. *et al.* Targeting isoprenylcysteine methylation ameliorates disease in a mouse model of progeria. *Science* **340**, 1330–1333 (2013).

- Court, H., Ahearn, I. M., Amoyel, M., Bach, E. A. & Philips, M. R. Regulation of NOTCH signaling by RAB7 and RAB8 requires carboxyl methylation by ICMT. *J. Cell Biol.* **216**, 4165–4182 (2017).
- Berndt, N., Hamilton, A. D. & Sebt, S. M. Targeting protein prenylation for cancer therapy. *Nat. Rev. Cancer* **11**, 775–791 (2011).
- Cox, A. D., Fesik, S. W., Kimmelman, A. C., Luo, J. & Der, C. J. Drugging the undruggable RAS: Mission possible? *Nat. Rev. Drug Discov.* **13**, 828–851 (2014).
- Lau, H. Y., Tang, J., Casey, P. J. & Wang, M. Isoprenylcysteine carboxylmethyltransferase is critical for malignant transformation and tumor maintenance by all RAS isoforms. *Oncogene* **36**, 3934–3942 (2017).
- Wang, T. *et al.* Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic RAS. *Cell* **168**, 890–903.e15 (2017).
- Wright, L. P. *et al.* Topology of mammalian isoprenylcysteine carboxyl methyltransferase determined in live cells with a fluorescent probe. *Mol. Cell Biol.* **29**, 1826–1833 (2009).
- Romano, J. D. & Michaelis, S. Topological and mutational analysis of *Saccharomyces cerevisiae* Ste14p, founding member of the isoprenylcysteine carboxyl methyltransferase family. *Mol. Biol. Cell* **12**, 1957–1971 (2001).
- Baron, R. A. & Casey, P. J. Analysis of the kinetic mechanism of recombinant human isoprenylcysteine carboxylmethyltransferase (ICMT). *BMC Biochem.* **5**, 19 (2004).
- Diver, M. M. & Long, S. B. Mutational analysis of the integral membrane methyltransferase isoprenylcysteine carboxyl methyltransferase (ICMT) reveals potential substrate binding sites. *J. Biol. Chem.* **289**, 26007–26020 (2014).
- Koide, A., Wojcik, J., Gilbreth, R. N., Hoey, R. J. & Koide, S. Teaching an old scaffold new tricks: monobodies constructed using alternative surfaces of the FN3 scaffold. *J. Mol. Biol.* **415**, 393–405 (2012).
- Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4**, 706–731 (2009).
- Lemmon, M. A., Treutlein, H. R., Adams, P. D., Brünger, A. T. & Engelman, D. M. A dimerization motif for transmembrane α -helices. *Nat. Struct. Biol.* **1**, 157–163 (1994).
- Yang, J. *et al.* Mechanism of isoprenylcysteine carboxyl methylation from the crystal structure of the integral membrane methyltransferase ICMT. *Mol. Cell* **44**, 997–1004 (2011).
- Long, S. B., Casey, P. J. & Beese, L. S. Reaction path of protein farnesyltransferase at atomic resolution. *Nature* **419**, 645–650 (2002).
- Li, X., Roberti, R. & Blobel, G. Structure of an integral membrane sterol reductase from *Methylomicrobium alcaliphilum*. *Nature* **517**, 104–107 (2015).
- Koide, S. Engineering of recombinant crystallization chaperones. *Curr. Opin. Struct. Biol.* **19**, 449–457 (2009).
- Tan, E. W., Pérez-Sala, D., Cañada, F. J. & Rando, R. R. Identifying the recognition unit for G protein methylation. *J. Biol. Chem.* **266**, 10719–10722 (1991).
- Pérez-Sala, D., Gilbert, B. A., Tan, E. W. & Rando, R. R. Prenylated protein methyltransferases do not distinguish between farnesylated and geranylgeranylated substrates. *Biochem. J.* **284**, 835–840 (1992).
- Anderson, J. L., Henriksen, B. S., Gibbs, R. A. & Hrycyna, C. A. The isoprenoid substrate specificity of isoprenylcysteine carboxylmethyltransferase: development of novel inhibitors. *J. Biol. Chem.* **280**, 29454–29461 (2005).
- Horowitz, S. *et al.* Conservation and functional importance of carbon-oxygen hydrogen bonding in AdoMet-dependent methyltransferases. *J. Am. Chem. Soc.* **135**, 15536–15548 (2013).
- Gordon, L. B., Rothman, F. G., López-Otín, C. & Misteli, T. Progeria: a paradigm for translational medicine. *Cell* **156**, 400–407 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Julius, C. D. Lima, M. Luo, N. P. Pavletich, R. K. Hite, S. Shuman and members of the Long laboratory for discussions. Beamlines 24-ID and 23-ID at the Advanced Photon Source are supported by NIH grants ACB-12002, AGM-12006, P41 GM103403 and S10 RR029205, under DOE contract DE-AC02-06CH11357. This work was supported, in part, by an American Heart Association Pre-doctoral Fellowship (M.M.D.), a Burroughs Wellcome Career Award (S.B.L.), the Geoffrey Beene Cancer Research Center at MSKCC (S.B.L.), a core-facilities support grant to MSKCC (P30 CA008748) and NIH grant U54-GM087519 (S.K.).

Author Contributions M.M.D. and L.P. cloned, expressed and purified proteins. A.K. and S.K. developed monobodies. M.M.D. performed all other experiments. M.M.D. and S.B.L. designed experiments, determined structures, analysed results and prepared the manuscript with contributions from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to S.B.L. (Longs@mskcc.org).

Reviewer Information Nature thanks M. Bergö, P. Casey and O. Nureki for their contribution to the peer review of this work.

METHODS

Cloning, expression, and purification of ICMT. *T. castaneum* (beetle) ICMT (UniProt accession D6WJ77) was selected as a candidate for protein purification and crystallization trials from among 76 eukaryotic ICMT orthologues that were evaluated using the fluorescence-detection size-exclusion chromatography (FSEC) pre-crystallization screening technique^{12,25}. The cDNA (synthesized by Bio Basic) was ligated into the EcoRI and SalI restriction sites of the *Pichia pastoris* expression vector pPICZ-C (Invitrogen Life Technologies) and encodes the full-length protein followed by a C-terminal antibody-affinity tag (Ala-Ala-Glu-Gly-Glu-Glu-Phe) that is recognized by the anti-tubulin antibody YL1/2²⁶. For the crystals of ICMT alone, two point mutations of surface residues were introduced to improve crystallizability (G151A and E154A). Transformation into *P. pastoris*, expression and cryo-lysis were performed as previously described²⁷.

Lysed cells (40 g) were re-suspended in 200 ml of buffer containing 10 mM Tris-HCl, pH 7.5, 150 mM KCl, 2 mM tri(2-carboxyethyl)phosphine (TCEP, Soltec Ventures; a 275 mM stock solution of TCEP was prepared in 1 M KOH to yield pH ~7.5), 2 mM CaCl₂, 25 μ M AdoHcy, 0.15 mg/ml DNase I (Sigma-Aldrich), 1:1,000 dilution of Protease Inhibitor Cocktail Set III (EDTA free, CalBiochem), 1 mM benzamidine (Sigma-Aldrich), 0.5 mM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSE, Gold Biotechnology) and 1:1,000 dilution of aprotinin (Sigma-Aldrich). Cell lysate was adjusted to pH 7.5 using 1 M KOH, 2 g decyl maltose neopentyl glycol (DMNG, Anatrace) was added to the cell lysate and the mixture was stirred at room temperature for 45 min to extract ICMT from the membranes. The sample was then centrifuged at 43,000g for 40 min at 12 °C and the supernatant was filtered through a 0.22- μ m polystyrene membrane (Millipore). YL1/2 antibody (IgG, expressed from hybridoma cells and purified by ion-exchange chromatography using standard methods) was coupled to CNBr-activated sepharose beads (GE Healthcare) according to the manufacturer's protocol. Approximately 0.4 ml of YL1/2 antibody beads were added to the sample for each 1 g of *P. pastoris* cells and the mixture was rotated at room temperature for 1 h. Beads were collected on a column, washed with four column volumes of buffer containing 10 mM Tris-HCl, pH 7.5, 150 mM KCl, 2 mM TCEP, 2 mM CaCl₂, 25 μ M AdoHcy and 1 mM DMNG, and the protein was eluted with buffer containing 100 mM Tris-HCl, pH 7.5, 150 mM KCl, 2 mM TCEP, 2 mM CaCl₂, 25 μ M AdoHcy, 1 mM DMNG and 5 mM Asp-Phe peptide (Sigma-Aldrich) or Glu-Glu-Phe peptide (Peptide 2.0).

ICMT-monombody co-crystallization in lipidic cubic phase. Following elution from the antibody-affinity column, the protein was combined with the monobody (designated MB-15) in a 1:3 molar ratio (ICMT:monobody). The mixture was concentrated to 500 μ l using a 30-kDa concentrator (Amicon Ultra, Millipore) and the ICMT-monombody complex was purified using a Superdex 200 Increase size-exclusion column (GE Healthcare) in 10 mM Tris-HCl, pH 7.5, 150 mM KCl, 5 mM TCEP, 2 mM CaCl₂, 25 μ M AdoHcy and 0.2 mM DMNG. The fractions corresponding to the ICMT-monombody complex and the free monobody were combined (to ensure an excess of monobody), 500 μ M AdoHcy was added, and the sample was concentrated to ~10 mg/ml using a 100-kDa Vivaspın-2 concentrator (Sartorius Stedim Biotech). Some of the excess monobody passed through the concentrator. Following concentration, 1 mM N-acetyl-S-geranylgeranyl-L-cysteine (AGGC, from a 100 mM stock in DMSO, Enzo Life Sciences) was added before crystallization.

The ICMT-monombody complex was combined with a mixture of 9.9 monoacylglycerol (monoolein, Nu-Chek Prep) at a ratio of 40:60 (v:v, ICMT:monoolein) using a manual syringe mixer at 20 °C following an established protocol¹⁴. A Gryphon LCP robot (Art Robbins Instruments) was used to dispense 50-nl mesophase drops onto 96-well glass sandwich plates (Marienfeld). The LCP boluses were overlaid with 800 nl precipitant solution. Drops were sealed with a glass coverslip, incubated at 20 °C and imaged periodically using a RockImager (Formulatrix). The precipitant solution for the best crystal consisted of 30% PEG 400, 100 mM NaCl and 100 mM Na HEPES, pH 7.5. Crystals also grew in other salts, including 100 mM LiSO₄ and 100 mM AmSO₄. Crystals appeared after one day and reached approximately 30–60 μ m in size within two to three days. To collect the LCP crystals, the glass coverslip was scored using a glass cutter, the LCP bolus was overlaid with ~2 μ l of additional precipitant solution, and the crystals were collected using MiTeGen MicroMounts and submerged in liquid nitrogen.

Crystallization in detergent. ICMT was purified as described above except that 1 mg/ml total brain lipids (Avanti) was added to the purification buffers. The elution from the antibody affinity column was concentrated to 500 μ l using a 50-kDa concentrator (Amicon Ultra, Millipore) and applied to a Superdex 200 Increase size-exclusion column (GE Healthcare) that was equilibrated in 10 mM Tris-HCl, pH 7.5, 150 mM KCl, 5 mM TCEP, 2 mM CaCl₂, 25 μ M AdoHcy, 1 mM DMNG and 0.02 mg/ml total brain lipids. Fractions containing ICMT were pooled, supplemented with 500 μ M AdoHcy, and concentrated to 5–10 mg/ml using a 50 kDa concentrator (Vivaspin-2; Sartorius). ICMT crystals were obtained

by vapour diffusion using 300 nl ICMT and 300 nl crystallization solution and a Mosquito crystallization robot (TTP Labtech) over reservoirs containing 100 μ l precipitant solution. The crystals were obtained from 24–28% PEG400 (v/v), 200 mM CaCl₂ or 200 mM MgCl₂, 50 mM Na acetate, pH 5.0–6.5, at 4 °C and reached a size of approximately 100–400 μ m within two to three weeks. The crystals were then transferred through a series of five steps using the components of an equilibrated drop to increase the PEG 400 concentration to 35% before flash-cooling in liquid nitrogen.

Monobody generation. General methods for phage- and yeast-display library sorting and gene shuffling have been described^{13,28}. His₁₀-tagged beetle ICMT protein, purified from *P. pastoris* and solubilized in 10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 25 μ M AdoHcy, 50 μ M AGGC and 1 mM DMNG, was mixed with an equimolar concentration of BTrisNTA, a high-affinity His-tag ligand conjugated with biotin²⁹, so as to capture the target proteins with streptavidin magnetic beads in phage-display selection and to detect them with fluorescent dye-labelled streptavidin in yeast-surface-display experiments³⁰. Two separate monobody libraries, denoted 'loop' and 'side', were used to generate monobodies with diverse binding-surface topography¹³. Four rounds of phage-display library sorting were performed using target concentrations of 100, 100, 100 and 50 nM for the first, second, third and fourth rounds, respectively, at room temperature. Three rounds of yeast display library sorting were performed, using a fluorescence-activated cell sorter (FACSaria, BD Biosciences). The first round was to enrich clones that did not bind to 500 nM BTrisNTA but without ICMT (negative sorting), and the second and third rounds were to enrich clones that bound to 500 nM and 250 nM ICMT-BTrisNTA complex (positive sorting). Target binding of individual clones was tested using yeast display and three were selected for purification. The monobody proteins were expressed in *Escherichia coli* using the expression vector pHBT, which adds an N-terminal His₆ tag followed by a biotin-acceptor tag and a TEV-cleavage site³¹. Co-crystallization with one of these monobodies (designated MB-15) yielded crystals with good diffraction. For purification of MB-15, frozen cells were mechanically lysed in a mixer mill (Retsch model PM 100; 8 \times 3 min at 400 r.p.m.) using steel balls. Lysed cells (6 g) were re-suspended in 50 ml buffer containing 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 0.2 mg/ml DNase I and 200 μ M AEBSE. The mixture was stirred at 4 °C for 30 min. The sample was then centrifuged at 43,000g for 30 min at 4 °C and the supernatant was filtered through a 0.22- μ m polystyrene membrane (Millipore). The supernatant was applied to a column containing 4 ml immobilized metal-affinity chromatography resin charged with cobalt (TALON, BD Biosciences), the resin was washed with 7.5 column volumes of buffer containing 20 mM Tris, pH 8.0 and 500 mM NaCl, and the protein was eluted with buffer containing 20 mM Tris, pH 8.0, 500 mM NaCl and 300 mM imidazole, pH 8.0. Following elution, 5 mM EDTA (pH 8.0) was added and the samples were centrifuged (43,000g, 10 min, 17 °C) to pellet any precipitated protein. The N-terminal affinity tag was removed by treatment with TEV protease (4 °C; 1:20, wt:wt, TEV:monobody) for 16 h and an additional 8 h with a 1:40 ratio of TEV:monobody). The amino sequence of MB-15 is: MKHHHHHHSSGLNDIF EAQKIEWHEENLYFQGSVSSVPTKLEVVAAAPTSLISWDAPAVTVVDLYVIT YGETGGNSPVQEFKVPKSKSTATISGLKPGVDYITITVYAFSSYYWPSYKGSPI SINRYT. The underlined portion indicates the peptide removed by TEV cleavage. Purified MB-15 was stored at –80 °C until use.

Structure determination. X-ray data were collected at beamlines 23ID-D and 24ID-C of the Advanced Photon Source (APS, Argonne). Diffraction data were collected using an oscillation angle of 0.3° and high redundancy was obtained by collecting data from multiple locations throughout the crystals. For phasing of the ICMT-monombody complex, Se-Met labelled ICMT protein was generated by producing the enzyme in High Five insect cells (Invitrogen) using a baculovirus system with standard techniques. Diffraction data were processed using the HKL3000 suite³². The crystals diffracted X-rays to 2.3 Å resolution and each asymmetric unit contained one ICMT-monombody complex. Initial experimental phases (35–4 Å) for the ICMT-monombody complex (space group P2₁2₁2₁) were determined using the SIRAS phasing method in SHARP and improved using solvent flattening³³. Five Se-Met sites were located; they correspond to the five methionine residues of ICMT. Electron density for the monobody was discontinuous in these maps. The monobody was placed in the asymmetric unit using molecular replacement (using an ensemble of monobody structures, PDB ID: 3UYO, 1FNA, 2QBW and 2OCF, and a homology model of MB-15 based on PDB ID: 3UYO) in Phaser³⁴. At this point, well-defined electron density was observed for the entire complex. The atomic model was constructed using Coot and improved through iterative cycles of refinement using CNS and PHENIX^{35–37}. Model validation was performed with MolProbity³⁸. Each unit cell contains one ICMT-monombody complex. Data collection, phasing, and refinement statistics are shown in Extended Data Table 1.

The structure of ICMT in the absence of the monobody (space group C222₁) was determined by molecular replacement using the refined 2.3 Å resolution structure of ICMT from the monobody complex as a search model in Phaser³⁴.

X-ray diffraction from the crystal of ICMT without monobody was anisotropic and the dataset was truncated to 4.2 Å resolution along b^* using the Diffraction Anisotropy Server³⁹. The electron density maps from Phaser showed continuous electron density throughout the model and indicated slight tilting of M5. The model was adjusted accordingly in Coot and refined in Phenix^{36,37}. A composite simulated-annealing omit-electron density map was constructed using Phenix³⁶ and this map confirmed that the only discernable difference in the overall structure was the tilting of M5 (Fig. 3e and Extended Data Fig. 3e). Molecular graphics figures were prepared using PyMOL (<http://www.pymol.org>).

Enzymatic assay. The catalytic activity of ICMT was measured using a previously described assay, with slight modifications, to monitor [³H]methyl incorporation into biotin-S-farnesyl-L-cysteine (BFC)^{11,12}. For these assays, ICMT was either purified (as described above; using 0.2 mM DMNG detergent) or present as a GFP-fusion protein in HEK293-cell lysate¹² (GFP was fused to the N terminus and an AAEGEF tag was present on the C terminus). For experiments using GFP-tagged ICMT, the preparation of the cell lysate and the determination of the ICMT concentration from the GFP fluorescence were as previously described¹². The enzyme, either as purified ICMT in DMNG detergent or as GFP-ICMT in HEK293-cell lysate, was diluted into reaction buffer containing 150 mM NaCl, 5 mM MgCl₂, 1 mM DTT and 100 mM HEPES pH 7.4 to yield an ICMT protein concentration of approximately 10 nM. For experiments using purified protein, the reaction buffer was supplemented with 10 μM DMNG. For the initial velocity curves, the data were fitted using GraphPad Prism 7 software as previously described¹². For experiments used to study the effect of the monobody on activity (Extended Data Fig. 2f, g), enzymatic assays were performed with varying concentrations of monobody, a fixed amount (~10 nM) of enzyme, 5 μM BFC and 4 μM AdoMet, using either purified beetle ICMT or one of the following GFP-tagged ICMT orthologues in HEK293 cell lysate: beetle ICMT, human ICMT or *Anopheles gambiae* ICMT¹².

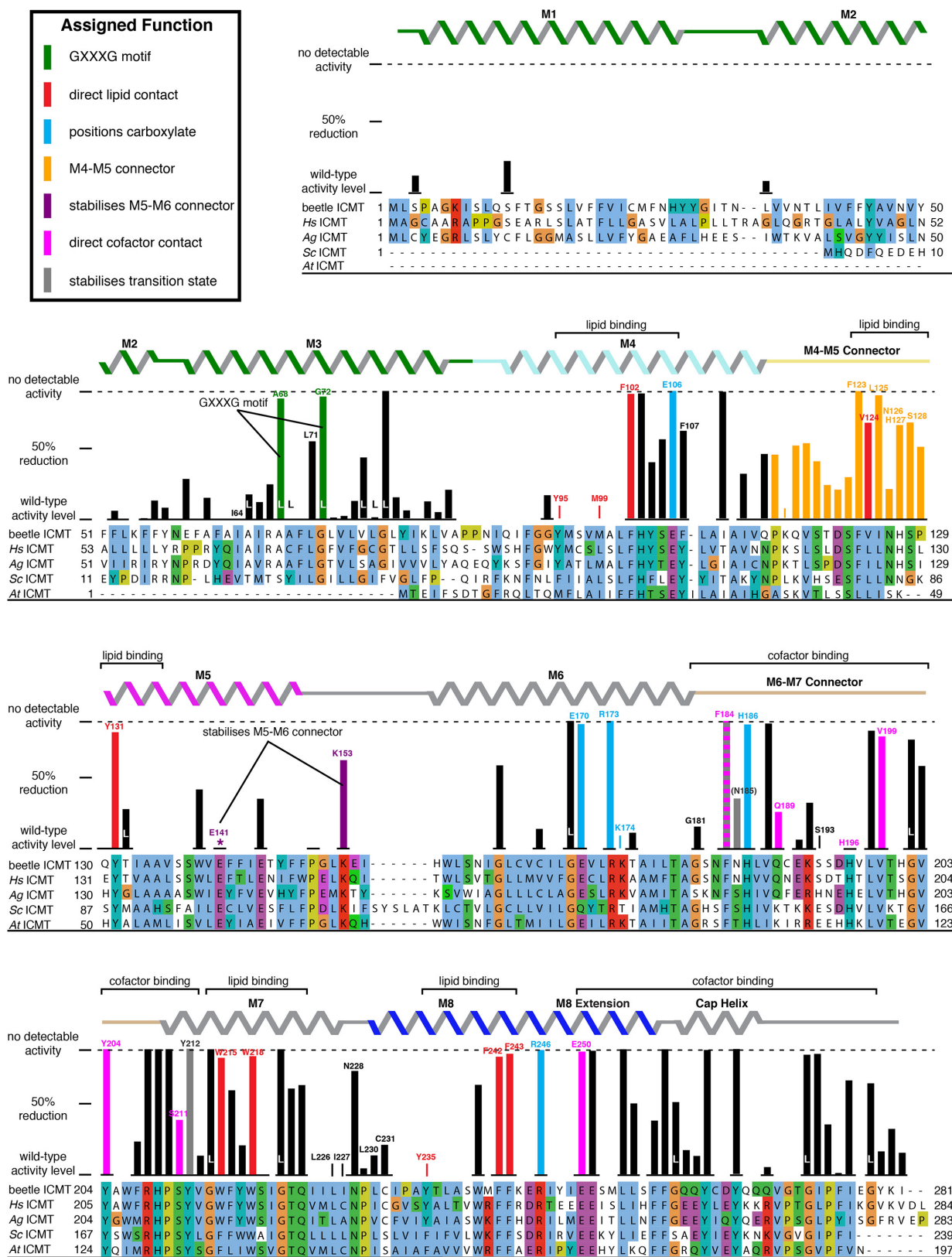
For initial velocity curves, the data were fitted to the following models using least-squares nonlinear regression with GraphPad Prism 7 software: (1) for the initial velocity versus AdoMet concentration curves, a Michaelis–Menten model was used, $Y = V_{\max}X/(K_m + X)$; (2) for the initial velocity versus BFC concentration curves when substrate inhibition was not observed, an allosteric-sigmoidal model was used, $Y = V_{\max}X^h/(K_{\text{half}}^h + X^h)$; and (3) for the initial velocity versus BFC concentration curves when substrate inhibition was observed, an allosteric-sigmoidal model that takes into account substrate inhibition was used, $Y = V_{\max}X^h/(K_{\text{half}}^h + X^h(1 + X^h/K_i^h))$. In these equations Y is the initial velocity, X is the substrate concentration, h is the Hill coefficient, V_{\max} is the maximum enzyme velocity, K_{half} and K_m are the concentrations of half-maximal velocity for sigmoidal and Michaelis–Menten models, respectively, and K_i is the inhibition constant. For experiments where the concentration of BFC was varied, 5 μM AdoMet was used. For experiments where the concentration of AdoMet was varied, 4 μM BFC was used. To account for the background activity due to endogenous human ICMT in the HEK293 cell lysate, initial velocity curves were constructed using lysate from cells transfected with GFP alone and subtracted for all analyses. The effective BFC concentration was corrected for the small amount of BFC that binds to plasticware used in the assay as described¹².

FSEC and western analysis of ICMT using a cleavable M2–M3 linker. cDNA encoding beetle ICMT was cloned into the pNGFP-EU expression vector²⁶ using the EcoRI and SalI restriction sites to create an ICMT construct with a 6 × His tag and GFP fused to the N terminus (His–GFP–ICMT). A PreScission protease (PS) cleavage site and several flanking amino acids were inserted into the loop between the M2 and M3 helices (introduced between Asn58 and Glu59 to yield the final sequence Asn58–SGSSGSLEVLFGQPSAGGSAGAAS–Glu59, where the underlined region is the PS site) using standard molecular biology techniques. These constructs, with or without the cleavage site, were expressed by transient transfection in HEK293 cells (~1.5 × 10⁶ cells using Lipofectamine 2000, Invitrogen). HEK293 cells were obtained from ATCC and were not tested

for mycoplasma contamination. The cells were pelleted (500g), 48 h after transfection and re-suspended in 300 μl of buffer consisting of 50 mM NaH₂PO₄, pH 7.5, 190 mM NaCl, 10 mM KCl, 20 mM DTT, 200 μM AdoHcy, 40 mM dodecyl maltoside (DDM, Anatrace) detergent and 1:500 dilution of Protease Inhibitor Cocktail Set III, EDTA-free (CalBiochem). Samples were rotated for 1 h at 4 °C to allow detergent extraction of ICMT from membranes, and then were centrifuged at 20,800g for 1 h at 4 °C. The supernatants, which contained detergent-solubilized ICMT, were divided into two portions. PreScission protease (2 μg; made in-house, but available from GE Healthcare) was added to one portion. The samples were incubated for 2 h at 4 °C and analysed by anti-His western blot (Anti-His₆ antibody, Roche Cat. # 04905318001) and by FSEC²⁶, using a Superose 6 column (GE Healthcare). The running buffer for FSEC was 20 mM NaH₂PO₄, pH 7.5, 190 mM NaCl, 10 mM KCl, 5 mM DTT and 1 mM DDM.

Data availability. Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession numbers 5V7P (ICMT–monobody) and 5VG9 (ICMT alone).

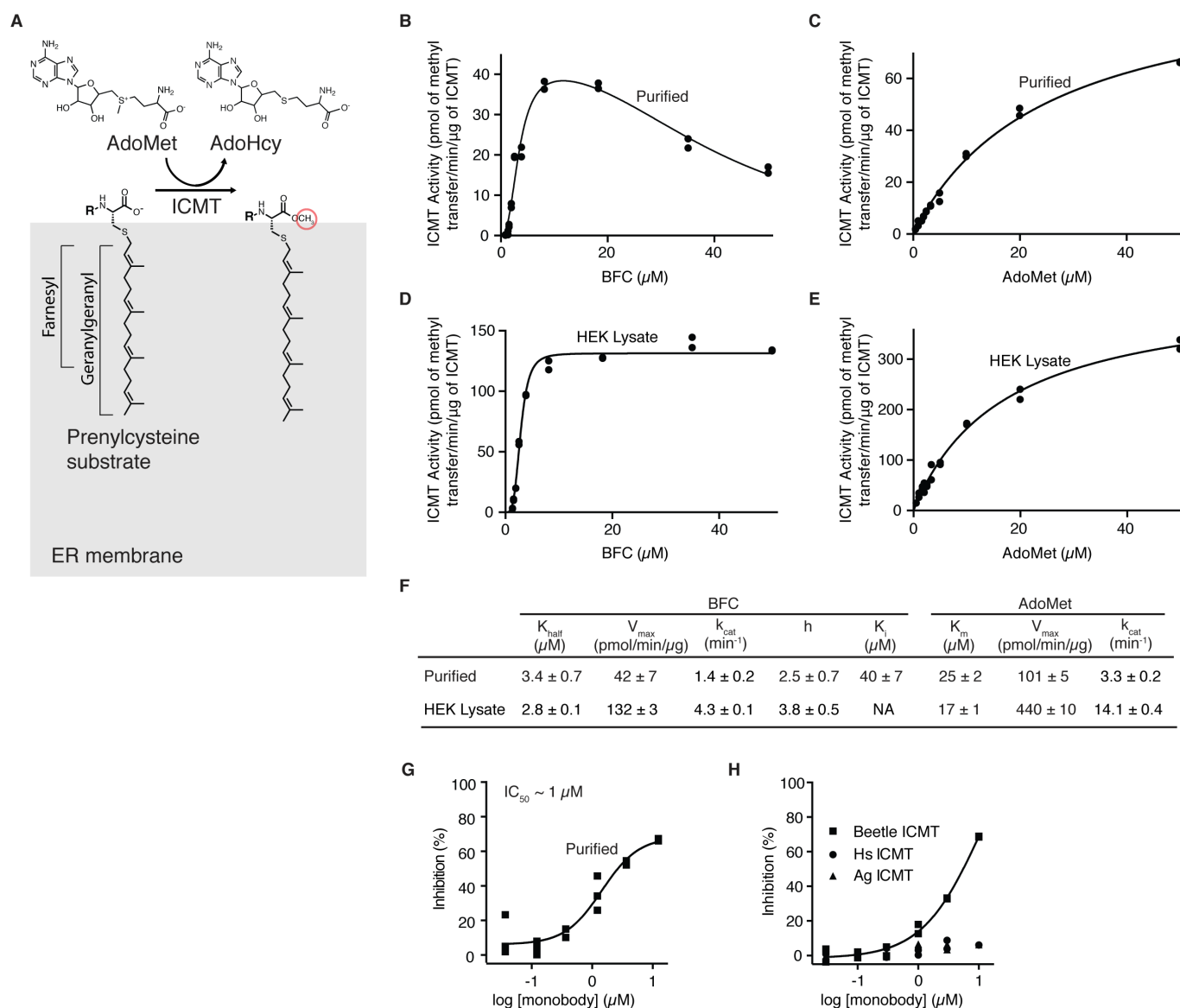
25. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
26. Killmartin, J. V., Wright, B. & Milstein, C. Rat monoclonal antitubulin antibodies derived by using a new nonsecreting rat cell line. *J. Cell Biol.* **93**, 576–582 (1982).
27. Long, S. B., Campbell, E. B. & Mackinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel. *Science* **309**, 897–903 (2005).
28. Koide, S., Koide, A. & Lipovšek, D. Target-binding proteins based on the 10th human fibronectin type III domain (¹⁰Fn3). *Methods Enzymol.* **503**, 135–156 (2012).
29. Koide, A. *et al.* Accelerating phage-display library selection by reversible and site-specific biotinylation. *Protein Eng. Des. Sel.* **22**, 685–690 (2009).
30. Stockbridge, R. B., Koide, A., Miller, C. & Koide, S. Proof of dual-topology architecture of Fluc F[−] channels with monobody blockers. *Nat. Commun.* **5**, 5120 (2014).
31. Sha, F. *et al.* Dissection of the BCR-ABL signaling network using highly specific monobody inhibitors to the SHP2 SH2 domains. *Proc. Natl Acad. Sci. USA* **110**, 14924–14929 (2013).
32. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62**, 859–866 (2006).
33. Bricogne, G., Vornrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
34. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
35. Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nat. Protoc.* **2**, 2728–2733 (2007).
36. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
37. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
38. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
39. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
40. Griggs, A. M., Hahne, K. & Hrycyna, C. A. Functional oligomerization of the *Saccharomyces cerevisiae* isoprenylcysteine carboxyl methyltransferase, Ste14p. *J. Biol. Chem.* **285**, 13380–13387 (2010).
41. Dharmalingam, S. *et al.* Structural basis of recognition of farnesylated and methylated KRAS4B by PDEδ. *Proc. Natl Acad. Sci. USA* **113**, E6766–E6775 (2016).
42. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).



Extended Data Figure 1 | See next page for caption.

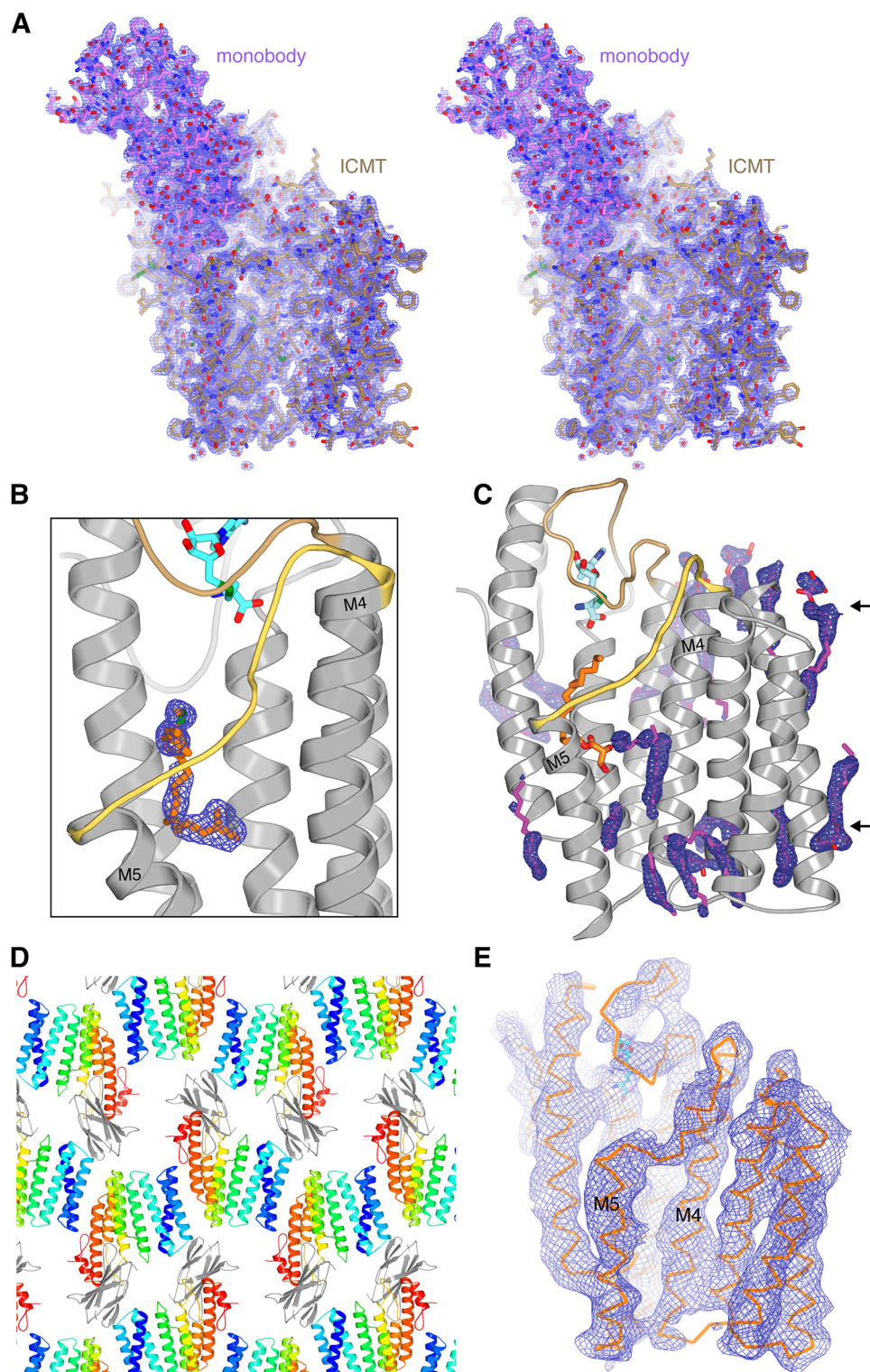
Extended Data Figure 1 | The structure identifies functions for key residues. ICMT is displayed as a progression from primary sequence alignment, to the effects of scanning mutagenesis (bar graphs), to observed secondary structure (α -helices depicted as ribbons), to the assigned role of amino acids. Results from scanning mutagenesis experiments¹² are plotted above the sequence alignment as a bar graph showing the reduction of specific activity in comparison to wild-type ICMT, with 0% representing wild-type activity and 100% reflecting no detectable activity (indicated by a horizontal dashed line). The inset key denotes the colouring of the bar graph according to the functional role of the amino acid inferred from the atomic structure. Labelled brackets above the secondary structure denote the general function of the indicated regions of the primary sequence. In the bar graph: magenta, amino acids that contact AdoHcy; red, amino acids that line the lipid-binding cavity; blue, arginine residues that are proposed to form hydrogen bonds with the carboxylate of the prenylcysteine substrate, and the residues that position them; grey, residues proposed to form hydrogen bonds to the methyl of AdoHcy in the transition state; hydrogen bonds made with backbone atoms are indicated

by parentheses surrounding the amino acid label. The mutagenesis data are derived from experiments using *A. gambiae* ICMT¹², and are normalized for expression level. Except where noted, the mutations were alanine substitutions. In some cases, leucine substitutions (L) were made (for example, when the wild-type amino acid was a glycine or alanine). The data represent triplicate measurements for each mutation and the mean s.d. is 11%. Gaps in the bar graph indicate amino acid positions that were not analysed by mutagenesis. Based on size-exclusion chromatography that was performed for each of the mutants, only the E141A mutation was found to be notably destabilizing (asterisk). A few mutations increased the activity relative to wild-type; these are shown as exhibiting 0% reduction in activity. The amino acid sequences included in the alignment are: *T. castaneum* ICMT (beetle ICMT), human (*Hs*), *A. gambiae* (*Ag*), *Saccharomyces cerevisiae* (*Sc*), and *Arabidopsis thaliana* (*At*) (UniProt accession numbers: D6WJ77, O60725, Q7PXA7, P32584 and Q93W54, respectively). The alignment is coloured according to the ClustalW convention.



Extended Data Figure 2 | Enzymatic activity and monobody inhibition of beetle ICMT. **a**, Schematic of the ICMT reaction. The shaded region represents the endoplasmic reticulum membrane; R, the protein portion of the substrate. In the minimal substrate AGGC, R is an acetyl group. For BFC, R is a biotin group. **b–e**, Activity of beetle ICMT, purified in DMNG detergent (**b**, **c**) or in cell lysate (**d**, **e**), shown as a plot of the formation of BFC-[^3H]methyl ester as a function of BFC concentration (**b**, **d**) or AdoMet concentration (**c**, **e**). For assays using HEK293 cells (**d**, **e**), ICMT was expressed as a fusion protein with GFP in order to quantify the amount of enzyme in the cell lysate (Methods). **f**, Kinetic parameters determined from the curves in **b–e**. Best-fit values (calculated in GraphPad Prism 7) are reported with the standard error of the fit. We observed a degree of substrate inhibition at higher concentrations

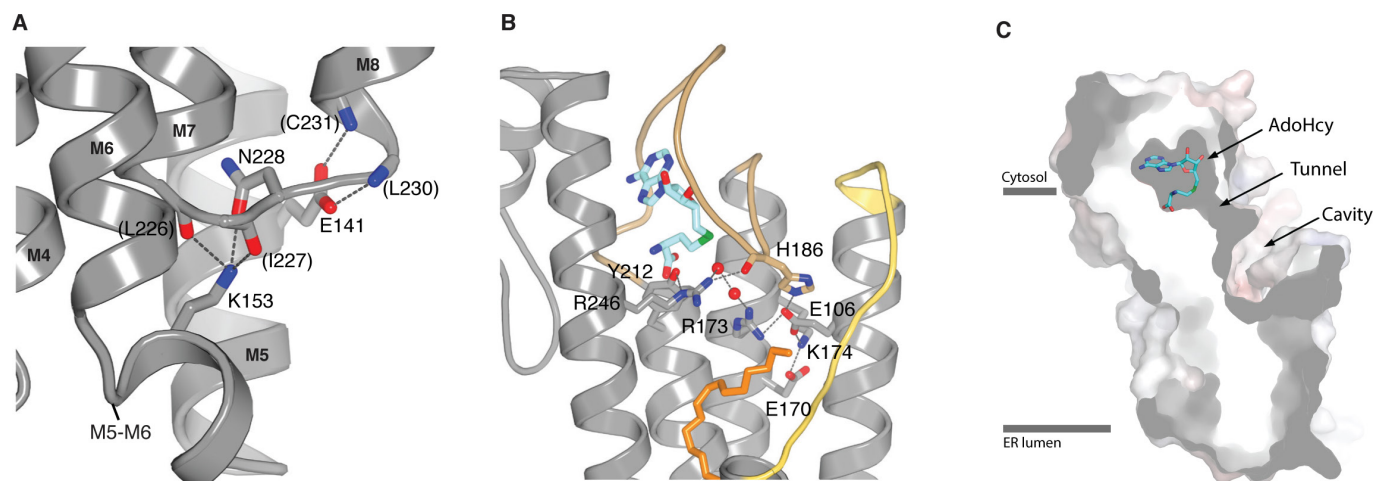
of BFC ($K_i \sim 40 \mu\text{M}$) when using ICMT in detergent (**b**, **f**), which may be due to the dispersive nature of BFC when it exceeds the detergent concentration ($10 \mu\text{M}$). **g**, Dose-response curve showing the inhibitory effects of the monobody on the activity of purified beetle ICMT in detergent. **h**, Comparison of the effects of the monobody on the activities of beetle, human, and *A. gambiae* ICMT in HEK293 cell lysates. GFP-ICMT fusion proteins were used for the three orthologues and the concentration of each enzyme was determined using the fluorescence of GFP (Methods). The monobody inhibited beetle ICMT with an IC_{50} of $\sim 7 \mu\text{M}$ in this assay, whereas no detectable inhibition of human or *A. gambiae* ICMT was observed. Individual data points are shown on the graphs (**b–e**, **g**, **h**).



Extended Data Figure 3 | Electron density maps and crystal lattice.

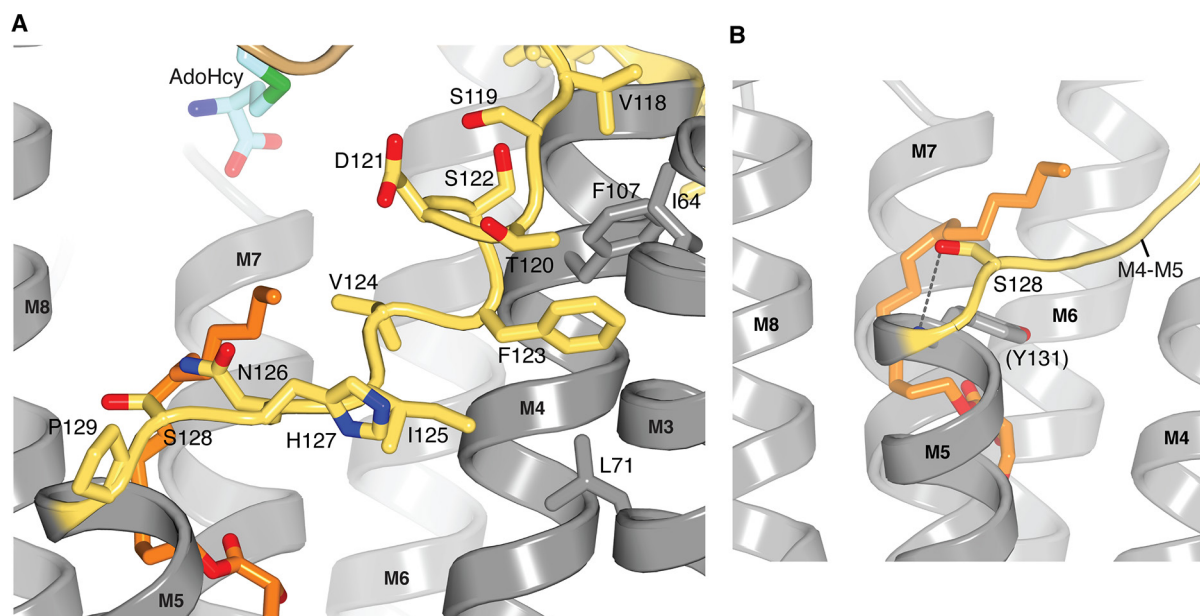
a, Stereo representation of the electron density for the ICMT-monombody complex (blue mesh, $2F_o - F_c$, contoured at 1.1σ , calculated from 35 to 2.3 Å resolution) drawn around the atomic model (stick representation). The monobody is coloured magenta; ICMT is tan. **b**, Electron density for the lipid in the active site from a composite simulated-annealing omit-electron density map ($2F_o - F_c$, contoured at 1σ and calculated from 35 to 2.3 Å resolution). The density would accommodate a geranylgeranyl lipid (orange sticks), as shown here. **c**, Electron density (blue mesh; $2F_o - F_c$ map contoured at 1σ) for ordered monoolein lipids (magenta sticks) around ICMT. Monoolein lipids surround the transmembrane region of ICMT

and collectively resemble a bilayer with typical thickness (arrows). The positioning of monoolein lipids in the vicinity of the M4-M5 connector (yellow) is consistent with the hypothesis that the enzyme would cause a slight depression in the membrane in this region, as illustrated in Extended Data Fig. 9. **d**, Crystal lattice in the ICMT-monombody complex, with ICMT coloured and the monobody in grey. **e**, A composite omit-electron-density map ($2F_o - F_c$, contoured at 1σ and calculated from 35 to 4.0 Å resolution) for the X-ray structure of ICMT without the monobody. The composite omit maps, which reduce model bias, were obtained using Phenix³⁶.



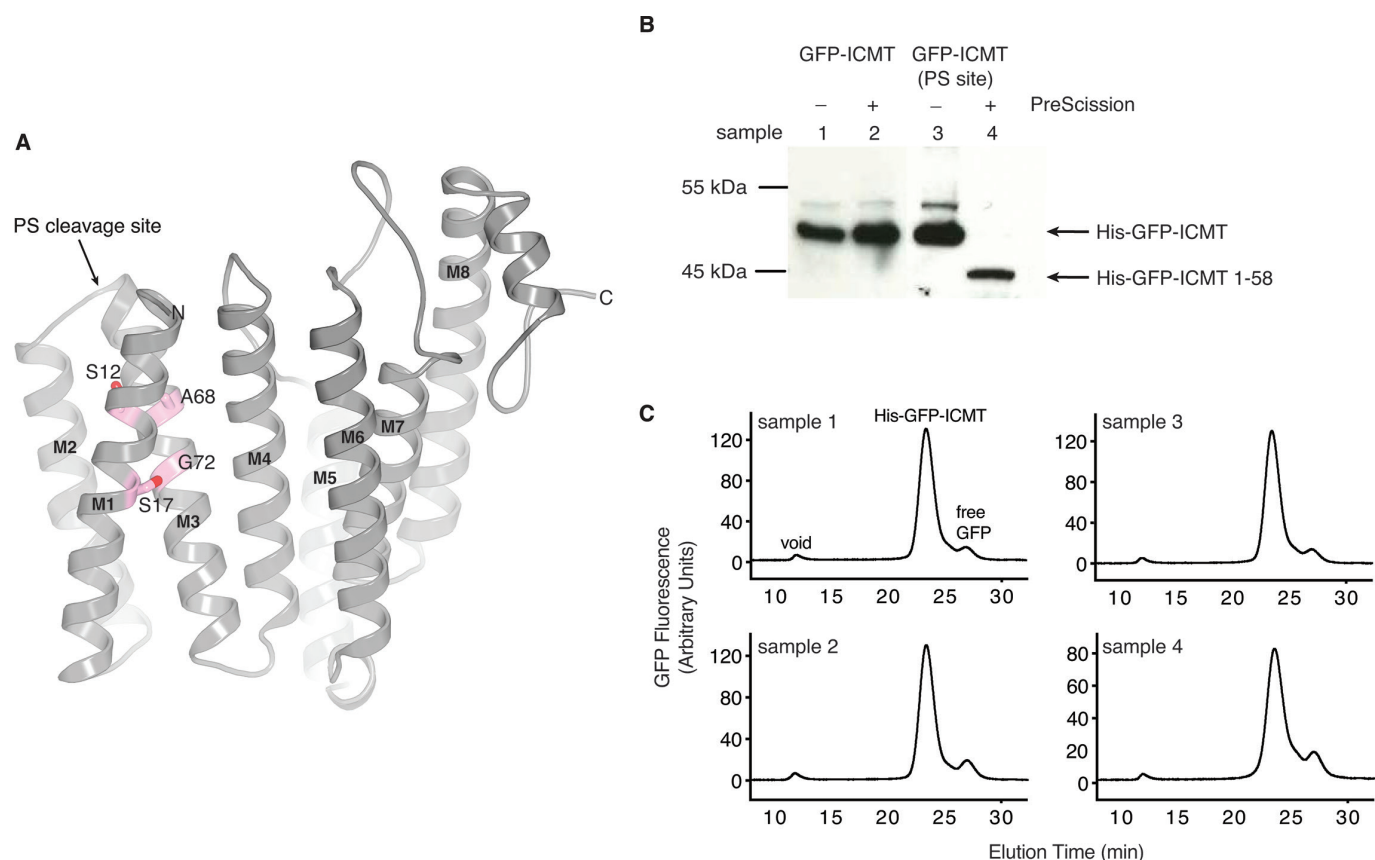
Extended Data Figure 4 | Interactions with the M5-M6 connector, interactions with active site arginine residues and cross-section of the active site. a, Interactions between the M5-M6 and M7-M8 connectors. The side chains of Glu141 (on M5) and Lys153 (on the M5-M6 connector) form hydrogen bonds (dashed lines) that cap the C-terminal end of M7 and the N-terminal end of M8, respectively. Portions of the amino acids involved are drawn as sticks. Parentheses indicate hydrogen bonds made with backbone atoms. Grey, carbon; blue, nitrogen; red, oxygen.

b, Hydrogen-bonding network involving Arg173 and Arg246. Bonds (dashed lines) are made with surrounding amino acids (labelled sticks) and two water molecules (red spheres) in the active site. **c**, Cutaway view of the molecular surface of ICMT (grey), showing labelled regions of the active site. The view is approximately orthogonal to that shown in Fig. 2a. AdoHcy is drawn as sticks; approximate boundaries of the membrane are indicated by horizontal bars.



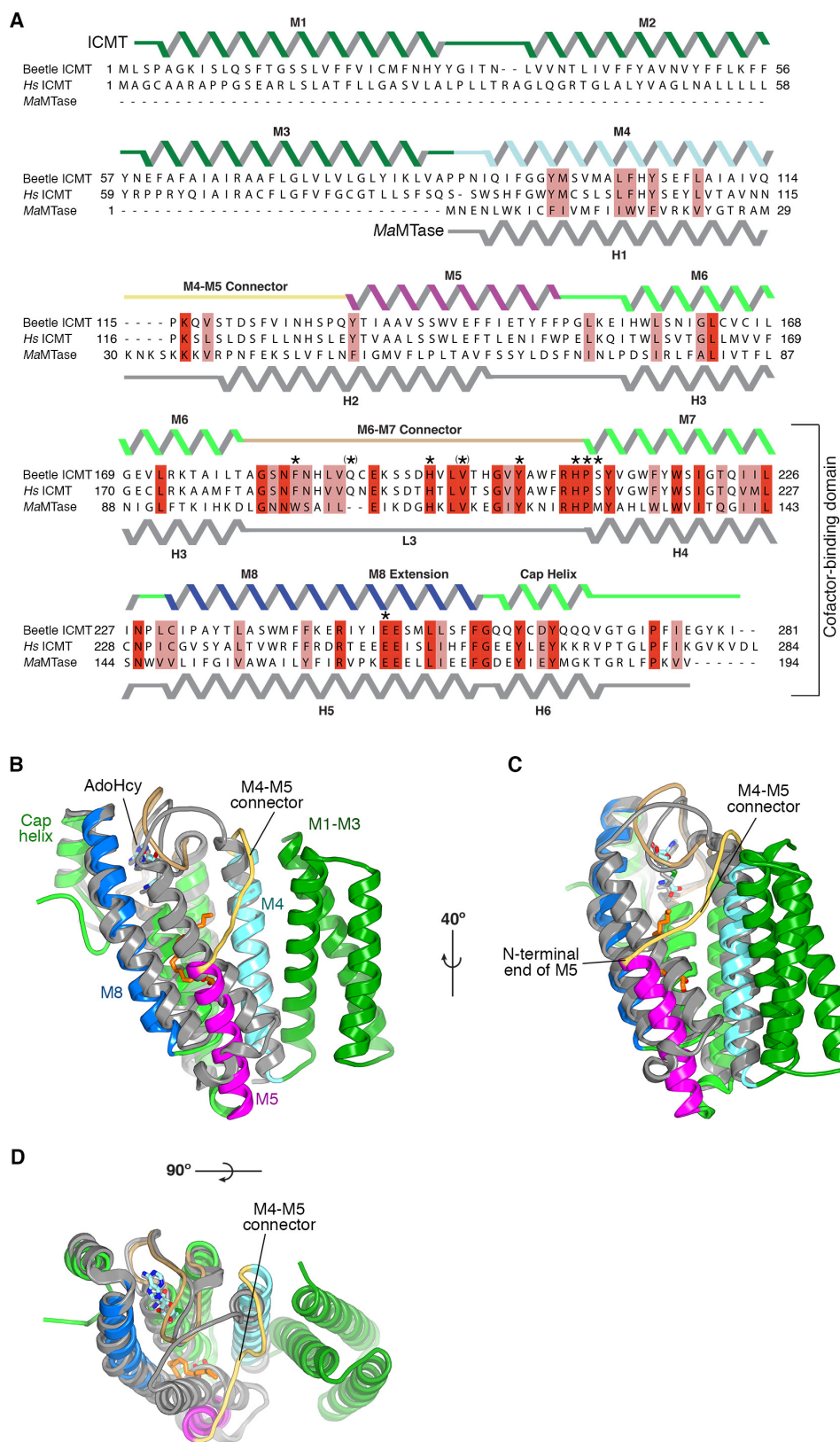
Extended Data Figure 5 | The M4–M5 connector. **a**, Close up view of the M4–M5 connector with side chains depicted as sticks. ICMT is predominately grey, with the M4–M5 connector coloured yellow. Amino acids that interact with Phe123 (Ile64, Leu71 and Phe107) are coloured

grey. **b**, The N-terminal end of the M5 helix is capped by Ser128. The side chain of Ser128 is shown in stick representation, with the hydrogen bond to the backbone nitrogen atom of Tyr131 indicated by a dotted line.



Extended Data Figure 6 | The M1–M2 portion of ICMT is an integral part of the enzyme. **a**, The GXXXG packing motif between helices M1 and M3 of ICMT. Residues on M1 and M3 that make contacts between these helices are drawn as sticks and coloured magenta. Because the M1 and M2 helices are not present in *S. cerevisiae* ICMT, we hypothesize that the GXXXG motif in the first transmembrane helix of yeast ICMT (equivalent to M3 of beetle ICMT) may cause dimerization of the yeast enzyme through packing of these helices⁴⁰, whereas ICMT enzymes that contain M1 and M2, which include human and beetle ICMT, are monomeric. The location of a PreScission protease cleavage site (PS site) that was introduced at Asn58, in the connection between the M2 and M3 helices, for experiments outlined in this figure, is indicated. **b**, Anti-His western blot showing that this cleavage site can be cleaved by PreScission protease. In this experiment, ICMT was expressed in HEK293 cells with an N-terminal His–GFP tag (His–GFP–ICMT), with or without the cleavage

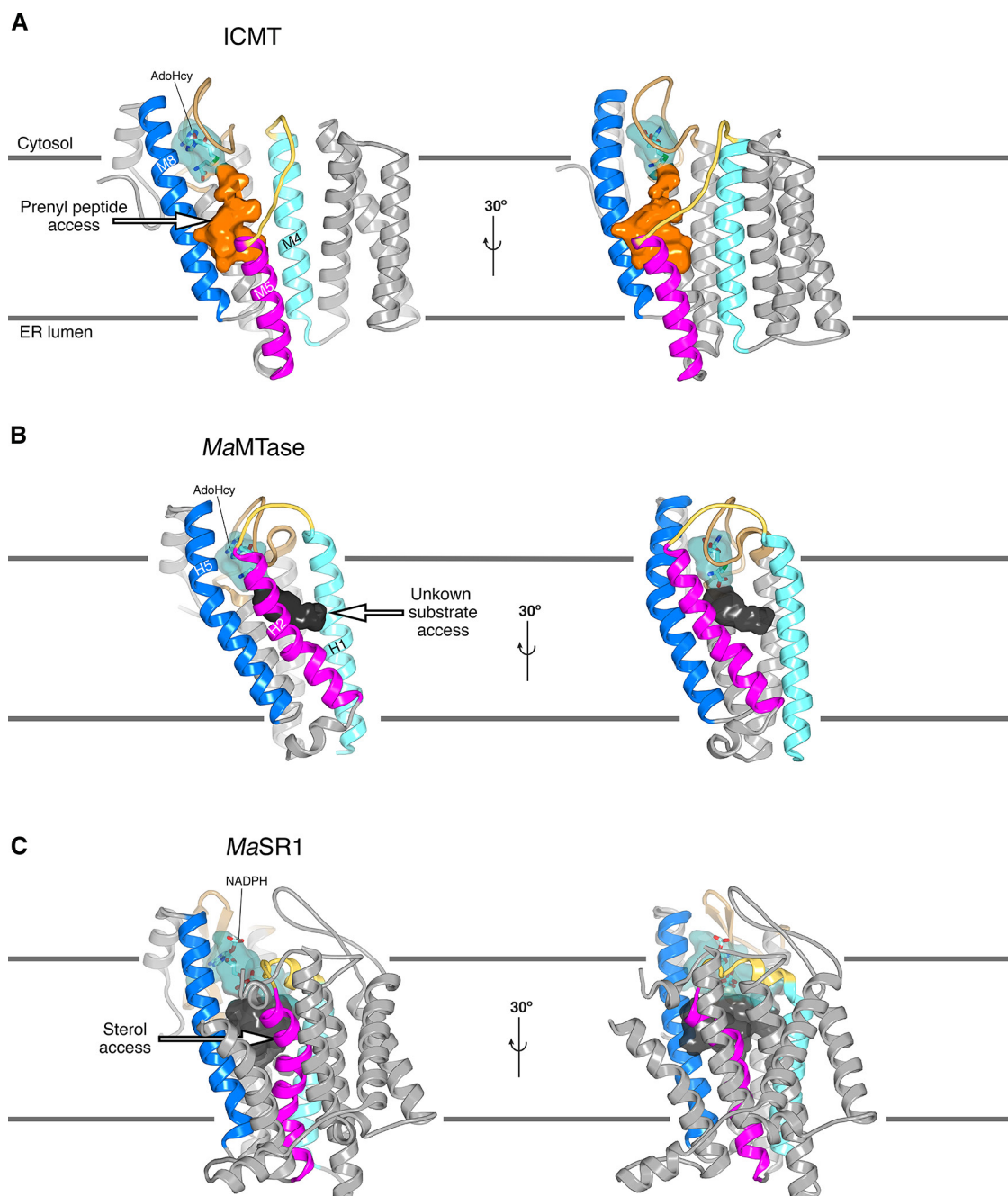
site. The addition of PreScission protease to detergent-solubilized sample 4 (His–GFP–ICMT with the cleavage site), but not control samples 1–3, results in a cleavage product consisting of His–GFP followed by amino acids 1–58 of ICMT (His–GFP–ICMT 1–58), which is detected by anti-His western blot. This confirms that the loop connecting the M2 and M3 helices is cut by the protease. Samples 1–3 are control experiments, as indicated. For gel source data, see Supplementary Fig. 1. **c**, FSEC profiles of the samples evaluated by western blot in **b** and numbered accordingly. Elution volumes for the void, His–GFP–ICMT and free GFP are indicated on the plot for sample 1. The cleavage of the M2–M3 loop by PreScission protease does not alter the elution profile (sample 4) in comparison to the other samples, which indicates that the cleaved portion (His–GFP–ICMT 1–58) is associated with the remainder of the enzyme via non-covalent interactions. No replicates of these experiments were performed.



Extended Data Figure 7 | Comparison of ICMT and MaMTase.

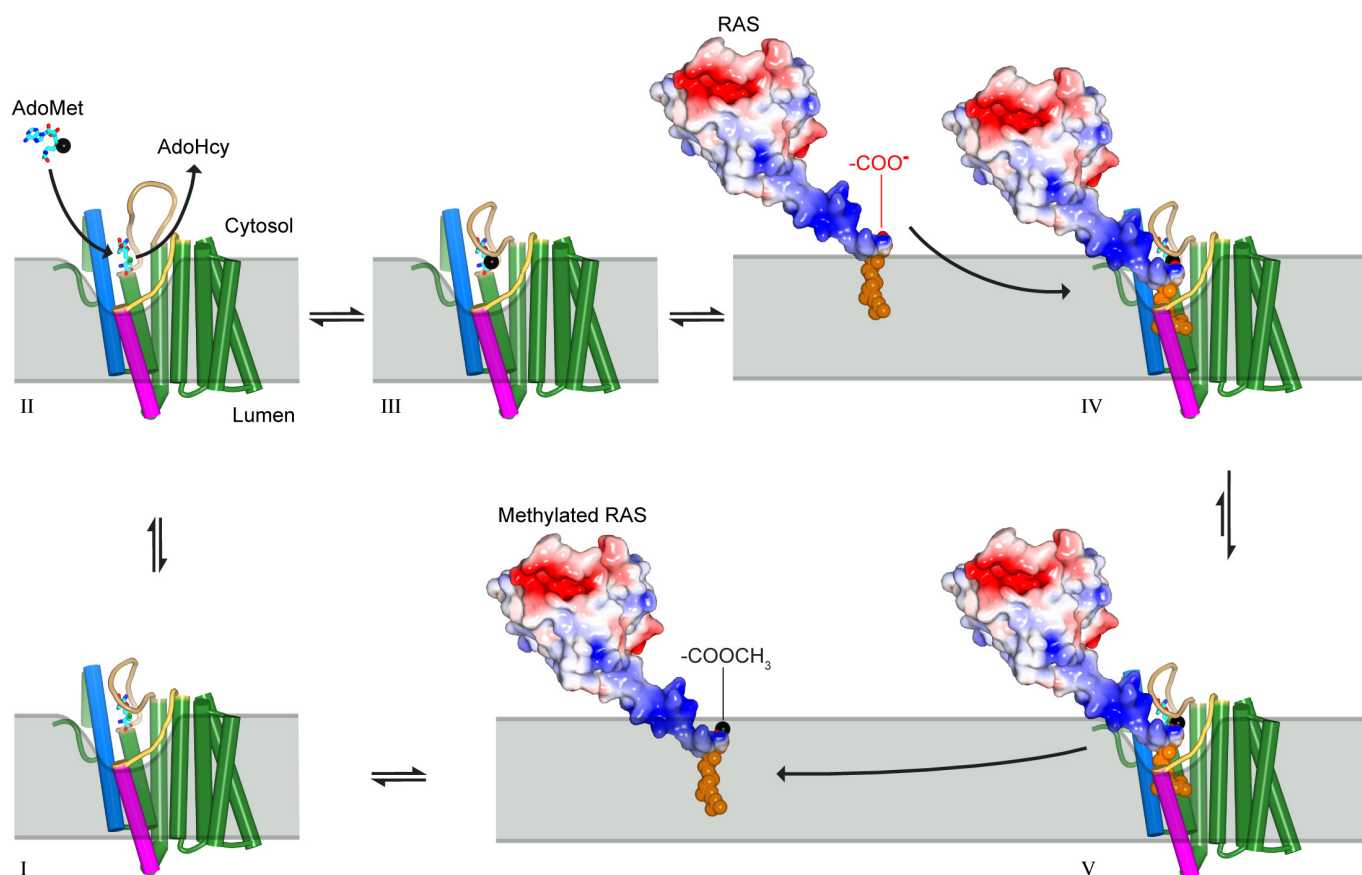
a, Sequence alignment between beetle ICMT, human ICMT, and MaMTase (referred to as Ma-ICMT in ref. 16). In the alignment, red colouring indicates identical residues and pink indicates similar residues. The secondary structures of ICMT and MaMTase (PDB ID: 4A2N) are shown above and below the alignment, respectively. The region of highest sequence conservation corresponds to the cofactor-binding domain

(bracketed); elsewhere there is low sequence similarity. Asterisks indicate amino acids of ICMT that contact AdoHcy; surrounding parentheses, where present, indicate that hydrogen bonds are made with backbone atoms. **b–d**, Superposition of the structures of beetle ICMT and MaMTase, shown from three vantage points in cartoon representations. ICMT is coloured according to the colouring of its secondary structure in **a**; MaMTase is coloured grey.



Extended Data Figure 8 | Comparison of the active sites of ICMT, MaMTase and MaSR1. **a**, Two orientations of the overall structure of ICMT are shown with the active site depicted as a molecular surface. The portion that binds AdoHcy is coloured cyan and is partially transparent to reveal a stick representation of the cofactor. The cavity that would be exposed to the membrane and forms the presumed binding site of the prenylcysteine moiety of the substrate is coloured orange. An arrow marks a pathway by which the prenyl group of the substrate could reach the active site via the membrane. Horizontal lines denote the approximate boundaries of the endoplasmic reticulum membrane. **b**, Overall structure of MaMTase (PDB ID: 4A2N) shown in two orientations, obtained by superposing with ICMT. A cavity, denoting the active site, is drawn as a molecular surface. The portion in which AdoHcy binds is coloured cyan; the remainder is dark grey and is likely to represent the binding site of its biological substrates, which are unknown. Structural elements are coloured as for ICMT in **a**. The cavity is exposed to the membrane

on the opposite side of the enzyme relative to ICMT. Instead of a crevice between the magenta- and dark-blue-coloured helices like ICMT (H2 and H5 of MaMTase, which roughly correspond to M5 and M8 of ICMT, respectively), MaMTase has a crevice between H1 (light blue) and H2 (magenta), suggesting that its substrates access it from the 'right' (arrow) rather than from the 'left'. In ICMT, this route is blocked by the M4–M5 linker. The dimensions of the cavity in MaMTase and its exposure to the membrane suggest that the substrates of MaMTase are hydrophobic molecules that are smaller than a farnesyl or geranylgeranyl prenyl group. **c**, Overall structure of the prokaryotic integral membrane sterol reductase MaSR1 (PDB ID: 4QUV)¹⁸. The orientation is based on superposition of the cofactor-binding domain with ICMT, with corresponding colouring. The carbon atoms of a bound NADPH cofactor are coloured cyan. A crevice between the magenta and dark blue-coloured helices may serve as access for lipophilic sterol substrates¹⁸ (arrow).



Extended Data Figure 9 | Proposed reaction cycle for ICMT. Major steps along the reaction coordinate (I–V). From the AdoHcy-bound state (I), a hinged displacement of the M6–M7 connector loop (II) allows release of AdoHcy and exchange for AdoMet from the cytosol (III). The C-terminal prenyl group of the substrate is located within the endoplasmic reticulum membrane before methylation, as depicted for KRAS4B in the figure, and approaches the enzyme by 2D diffusion within the membrane. The prenylcysteine reaches the active site by passing between the M5 and M8 helices (coloured magenta and blue, respectively). In the ternary complex (IV), substrate–enzyme contacts are limited to interactions with the prenyl group and the C-terminal carboxylate, giving rise to specificity for these elements within the context of a wide range of protein substrates. A cytosolic cleft above the M5–M8 crevice that leads to the active site accommodates the polar C-terminal peptide of

protein substrates (for example, the polybasic region of KRAS4B, coloured blue). The inhibitory monobody occupies this region. Catalysis proceeds from the ternary complex, and the product, made more lipophilic by methylation and neutralization of a negative charge, is able to diffuse into the membrane (V). ICMT is shown as a cartoon with helices drawn as cylinders and the AdoHcy or AdoMet cofactors depicted as cyan sticks. A black sphere indicates the methyl group of AdoMet. Prenylated KRAS4B (based on PDB ID: 5TAR⁴¹) is depicted as a molecular surface and coloured according to electrostatic potential (red, negative; blue, positive) with its prenyl group shown as orange spheres. The endoplasmic reticulum membrane is depicted in grey and curved in the vicinity of the active site to suggest that the enzyme might modulate the local distribution of lipid molecules to facilitate substrate access.

Extended Data Table 1 | Data collection, phasing and refinement statistics

	ICMT-monobody Native	ICMT-monobody SeMet	ICMT Native
Data Collection	APS 24-ID-C	APS 23-ID-D	APS 23-ID-D
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	C222 ₁
Wavelength (Å)	0.9790	0.9791	1.0332
Cell dimensions:			
a, b, c (Å)	40.6, 87.7, 147.7	40.8, 88.7, 148.4	51.6, 123.6, 236.1
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	31 - 2.3 (2.34 - 2.3)	40 - 3.0 (3.05 - 3.0)	33 - 4.0 (4.07 - 4.0)
No. of crystals	1	9	1
R _{merge} (%)	20.6 (97.9)	64.2 (>100.0)	7.1 (>100.0)
R _{dim} (%)	6.2 (29.5)	9.6 (57.0)	1.2 (34.0)
I/σI	13.3 (2.5)	15.5 (2.0)	85 (3.0)
Completeness (%)	97.2 (99.8)	100.0 (100.0)	100.0 (100.0)
Redundancy	11.2 (11.1)	45.7 (47.9)	34.6 (35.6)
SIRAS Phasing		30 - 4.0 Å	
No. of sites		5	
Phasing Power (iso / ano)		0.129 / 0.292	
R _{cullis} (iso / ano) (%)		95.5 / 97.5	
Figure of Merit (acentric / centric)		0.183 / 0.047	
Refinement			
Resolution (Å)	31 - 2.3 (2.4 - 2.3)		33 - 4.0 (5.0 - 4.0)
No. of reflections	23,203		6,119
R _{work} (%)	21.4 (26.9)		36.9 (45.0)
R _{free} (%)	24.6 (32.2)		38.9 (42.2)
No. atoms			
Protein	3039		2250
Ligands	319		26
Water	92		
Average B-factors (Å ²)			
Protein	33.4		209.6
Ligands	58.6		200.7
Water	38.7		
Ramachandran (%)			
Favoured	98.1		97.3
Outliers	0.0		0.0
R.m.s. deviations			
Bond lengths (Å)	0.003		0.004
Bond angles (°)	0.595		0.745
Clash Score	2.3		24

Data collection statistics are from HKL3000 (ref. 32); phasing statistics are from SHARP³³; refinement statistics are from PHENIX³⁶. CC1/2 is defined in ref. 42; Clash Score is defined in ref. 38. $R_{\text{sym}} = \sum |I_i - \langle I_i \rangle| / \sum I_i$, where $\langle I_i \rangle$ is the average intensity of symmetry-equivalent reflections. Phasing power = r.m.s. ($|F|/\epsilon$), where $|F|$ is the heavy-atom structure-factor amplitude and ϵ is the residual lack of closure error (r.m.s. is root mean square). R_{cullis} is the mean-residual lack-of-closure error divided by the dispersive or anomalous difference. R -factor = $\sum |F_o - F_c| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. R_{free} = R -factor calculated using a subset (~5%) of reflection data chosen randomly and omitted throughout refinement. Numbers in parentheses indicate the highest-resolution shells and their statistics.

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature25017

Corrigendum: Reductions in global biodiversity loss predicted from conservation spending

Anthony Waldron, Daniel C. Miller, Dave Redding,
Arne Mooers, Tyler S. Kuhn, Nate Nibbelink,
J. Timmons Roberts, Joseph A. Tobias & John L. Gittleman

Nature **551**, 364–367 (2017); doi:10.1038/nature24295

In the Abstract of this Letter, the following sentence: ‘Here we demonstrate such a model, and empirically quantify how conservation investment between 1996 and 2008 reduced biodiversity loss in 109 countries (signatories to the Convention on Biological Diversity and Sustainable Development Goals), by a median average of 29% per country’ should have read: ‘Here we demonstrate such a model, and empirically quantify how conservation investment reduced biodiversity loss in 109 countries (signatories to the Convention on Biological Diversity and Sustainable Development Goals), by a median average of 29% per country between 1996 and 2008’. This has been corrected online.

ERRATUM

doi:10.1038/nature24663

Erratum: Observation of the hyperfine spectrum of antihydrogen

M. Ahmadi, B. X. R. Alves, C. J. Baker, W. Bertsche,
E. Butler, A. Capra, C. Carruth, C. L. Cesar, M. Charlton,
S. Cohen, R. Collister, S. Eriksson, A. Evans, N. Evetts,
J. Fajans, T. Friesen, M. C. Fujiwara, D. R. Gill, A. Gutierrez,
J. S. Hangst, W. N. Hardy, M. E. Hayden, C. A. Isaac, A. Ishida,
M. A. Johnson, S. A. Jones, S. Jonsell, L. Kurchaninov,
N. Madsen, M. Mathers, D. Maxwell, J. T. K. McKenna,
S. Menary, J. M. Michan, T. Momose, J. J. Munich,
P. Nolan, K. Olchanski, A. Olin, P. Pusa, C. Ø. Rasmussen,
F. Robicheaux, R. L. Sacramento, M. Sameed, E. Sarid,
D. M. Silveira, S. Stracka, G. Stutter, C. So, T. D. Tharp,
J. E. Thompson, R. I. Thompson, D. P. van der Werf &
J. S. Wurtele

Nature **548**, 66–69 (2017); doi:10.1038/nature23446

Owing to a technical error, author J. S. Wurtele was listed incorrectly as a corresponding author instead of author J. S. Hangst in the HTML version of this Letter (the PDF version was correct). This has been corrected online.

CORRECTIONS & AMENDMENTS

ERRATUM

doi:10.1038/nature25160

Erratum: Genetic effects on gene expression across human tissues

GTEEx Consortium

Nature **550**, 204–213 (2017); doi:10.1038/nature24277

In the HTML version of this Article, the GTEEx Consortium author list shown underneath the title was incorrectly displayed as: ‘GTEEx Consortium, Lead analysts:, Laboratory, Data Analysis & Coordinating Center (LDACC):, NIH program management:, Biospecimen collection:, Pathology:, eQTL manuscript working group:, Alexis Battle, Christopher D. Brown, Barbara E. Engelhardt & Stephen B. Montgomery’ instead of ‘GTEEx Consortium’ (with a link to the complete author list at the bottom of the page). This has been corrected online.

TECHNOLOGY FEATURE

TECHNOLOGY TO WATCH IN 2018

Thought leaders reveal the technologies and topics likely to transform life–science research in the year ahead.

CHOMBOSAN/ALAMY



The Internet of Things has transformed many aspects of our lives and is now, along with other breakout technologies, poised to transform life-science research.

GEORGE CHURCH Recoding the genome

Geneticist, Harvard Medical School, Boston, Massachusetts.

For all the excitement surrounding the gene-editing tool CRISPR, it is not that efficient or precise. It's hard to make many changes at once. My lab has set the record so far — making 62 modifications to the genome of a single cell — but we have compelling applications that need a greater number of simultaneous changes. Now, however, we have the technologies required to make this feasible.

'Codon recoding' is a completely generic way to make any organism resistant to most or all viruses and requires tens of thousands of precise changes per cell. Each codon, a section of DNA three bases long, such as TTG, corresponds to a specific amino acid, such as leucine, or a translational signal (start, stop and so on). Given that there are six codons for the leucine, we can switch any one for another, taking advantage of the redundancy built into the genetic code. Once done with those swaps, we delete the gene for the leucine transfer RNA (tRNA) that matches up with the swapped-out codons, so the cell can no longer recognize that sequence.

Now, when a virus infects a cell that has all of these codons recoded, it cannot translate its proteins from its messenger RNA because of the missing tRNA, and the virus will die.

Viruses are not that robust; it doesn't take much to throw them out of whack.

To make multiple, precise changes at once, we use the multiplexed automated genome engineering (MAGE) technique. Short segments of genetic material containing the precise base-pair changes you want to make are introduced into cells that are prevented from making DNA-mismatch repairs. After a few rounds of cellular replication the changes are incorporated fully into the bacterial genome.

Theoretically, this can be done in every organism for which viruses are a problem — microorganisms used in the dairy industry and agriculturally important plants and animals. In addition, researchers could make virus-resistant pigs whose organs can be used for transplants, and virus-resistant ►

► human cells to use for producing pharmaceuticals and vaccines.

What is really gee whiz here is that you have the potential to make an organism resistant to all viruses — even viruses that have never been studied. But there are many other things that recoding can accomplish. Pamela Silver at Harvard Medical School and Daniel Gibson at Synthetic Genomics in La Jolla, California, have collaborated to develop another recoding technology to improve vaccine strains of *Salmonella typhimurium*.

Researchers could also recode an organism to incorporate non-standard amino acids in proteins to enable chemistries that don't exist in current organisms: amino acids that fluoresce, resemble nucleic acids or form unusual bonds. Whole new dimensions of biochemistry emerge when you are not limited to the universal and ordinary 20 amino acids. Jason Chin's lab at the MRC Laboratory of Molecular Biology in Cambridge, UK, is using this approach to make precise alterations at the molecular level of well-known proteins in fruit flies.

Last, but not least, recoding provides a potent strategy for bio-containment. If a virus-resistant organism were to escape, even if they weren't 'bad' for the environment, they would take over natural niches and 'win'. Using one of these non-standard amino acids, you can engineer an organism that can grow only if it is given that certain nutrient. The result is an 'escape-proof' strategy for experimental organisms used in the laboratory.

XIAOWEI ZHUANG Mapping the transcriptome

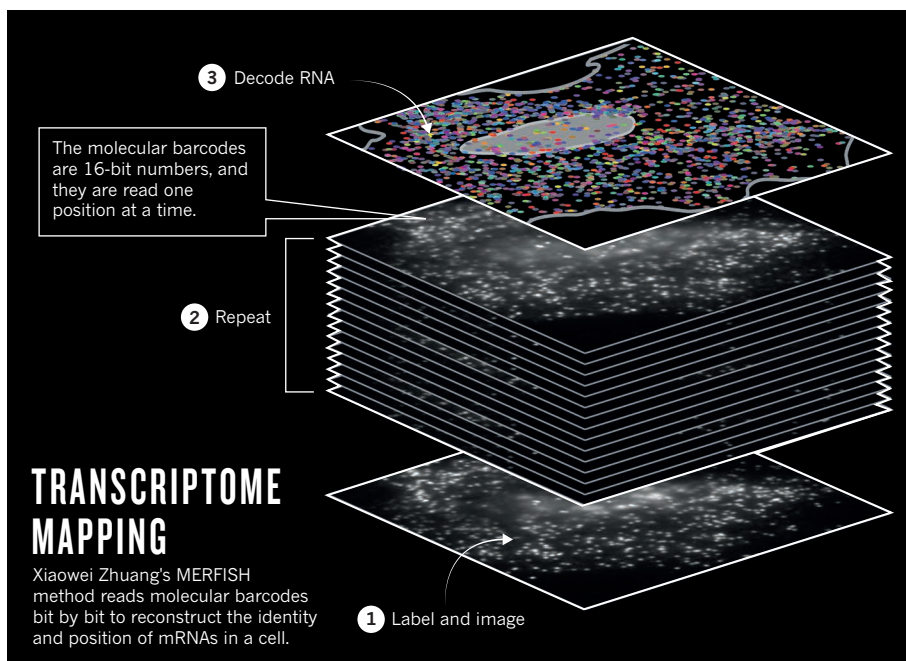
Director, Center for Advanced Imaging, Harvard University, Cambridge, Massachusetts.

A new global initiative to identify all cell types in the human body and map their spatial organization — the recently launched Human Cell Atlas (HCA) initiative — is a grand goal. A project of this scale will need many complementary technologies.

Single-cell RNA sequencing is a powerful way to identify different cell types and an important tool for creating the HCA, but it requires taking a tissue apart into individual cells and then isolating the RNA. What's not preserved is the spatial context of cells in a tissue — how they are organized and interact.

We'd like a technology that can provide this spatial context by imaging the transcription profiles of cells in intact tissue. My lab is developing MERFISH, or multiplexed error-robust fluorescence *in situ* hybridization, an image-based, single-cell transcriptomics approach.

MERFISH uses error-robust barcodes



to identify each different type of RNA in the cell, and combinatorial labelling and sequential imaging to detect these barcodes in a massively multiplexed manner (see 'Transcriptome mapping').

We've already demonstrated the ability to image 1,000 different messenger RNAs (mRNAs) in single cells. With further development, MERFISH has the potential to detect the whole transcriptome in cells in intact tissues.

This spatially resolved RNA-profiling data will give us a physical picture for the HCA — we can image individual cells, categorize them by their gene-expression profiles and map their spatial organization. It can be combined with data on the morphology and function of cells obtained by other imaging technologies to further enrich that picture.

At the moment, our picture of the cell atlas is mostly incomplete. If you don't have a global picture, you just don't know what you are missing — let alone how to design effective therapeutics to intervene in the case of disease.

ELAINE MARDIS Boosting cancer vaccines

Co-executive director, Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, Ohio.

In the field of cancer immunogenomics, researchers want to know which of the mutated proteins encoded by the cancer genome are capable of eliciting an immune response in a given individual. Such proteins, called neo-antigens, could be utilized to

develop personalized cancer vaccines or indicate other treatments.

One exciting technology that could be used to study these neo-antigens is CyTOF, a so-called mass-cytometry method for identifying cells that express specific proteins.

In typical flow cytometry, researchers mix antibodies labelled with a fluorescent molecule with cells to tag proteins of interest. Then the cells are analysed, one by one, to measure their relative abundances on the basis of those proteins. CyTOF replaces the limited handful of fluorescent tags with metallic labels that are detected in a mass spectrometer — 100 or more different labels at once, compared with perhaps a dozen in the case of flow cytometry.

This technology could transform the field of cancer immunogenomics, by enabling researchers to work out which neo-antigens produced by an individual's cancerous cells are the most abundant and most likely to elicit a strong reaction from the immune system. Researchers could then use that information to create personalized anti-cancer 'vaccines'. These, used in combination with new cancer drugs that release the brakes on the immune system, could put people with cancer in a position to fight off their own disease.

But for any given neo-antigen predicted by the genome, it's guesswork as to whether it will elicit a significant immune response. CyTOF gives us insight into that question, by letting us quantify the binding strengths of multiple predicted peptides to the person's T cells.

And it's not just for cancer genomics. CyTOF can be used to track the abundance and make-up of any suite of proteins produced by cells, as long as you can find antibodies to bind your proteins of interest. It's allowing us to ask questions at the protein level in a much more multidimensional and precise way than before.

RUEDI AEBERSOLD

Linking genotype and phenotype

Systems biologist, Institute of Molecular Systems Biology, ETH Zurich, Switzerland.

Clearly, we're living in a very interesting time — there is an enormous amount of high-quality genomic information on genetic variability. At the same time, we can collect masses of health-related data on the human population, ranging from the number of steps taken in a day to blood pressure and clinical imaging. The trick is to relate the two to each other. Especially in medicine, if we want to translate a genetic variation into a treatment, then we need mechanistic insights into the processes that are disrupted by disease.

The key to this link is the analysis of protein complexes, which are the functional units of cells. How do we go from big data — such as the genome of an ovarian tumour — to working out which protein complexes are perturbed and how?

One path blends computation and quantitative proteomics, in which several thousand

proteins are consistently and accurately quantified across cohorts of tumour and control samples. Such data sets can now be generated using mass-spectrometry techniques such as SWATH-MS (sequential window acquisition of all theoretical mass spectra). Complexed proteins would be expected to have a high degree of co-variation — that is, to increase or decrease in abundance synchronously. But if the complex is perturbed and loses subunits owing to mutation or structural changes, the subunit co-variance would be different. This is one way to pinpoint protein complexes that are perturbed in cancer.

Such altered complexes can then be studied at the structural level using cryo-electron microscopy single-particle analysis or cryo-electron tomography (CET), both of which can image molecules at a resolution of around 5–10 ångströms. This is high enough to visualize how mutations change the composition, topology and structure, and by inference the function, of the affected complex.

CET also has the power to reveal how structures vary with other modulations, such as the addition of a phosphate group to the completed molecule. A really significant advance for 2018 is the refinement of focused-ion beam milling. This technique takes a mammalian cell or tissue section

that is otherwise too thick for CET and mills out a thin window of the cell, such that the structure of a particular protein complex can be observed in the context of the cell.

Together, these technologies will increase our understanding of how a protein complex is perturbed at the molecular level in disease. And they will illuminate how one would engineer a drug to destroy it, inactivate it or restore its normal activity.

REBECCA CALISI RODRÍGUEZ

Extending genome sequence analysis

Reproductive biologist, University of California, Davis.

When I was entering graduate school, I was fascinated by the discovery in 2000 of a completely new hormone, gonadotropin inhibitory hormone (GnIH), that inhibits the reproductive axis when animals are stressed. Studies of GnIH are completely changing our understanding of how the brain regulates reproduction. I thought, “Hell, what else don't we know about? When will the

next discovery happen that completely changes the way we understand reproduction?”

Today, thanks to high-throughput DNA sequencing of genomes and transcriptomes, the rate of discovery is rising sharply. It took around US\$3 billion to sequence the human genome 15 years ago. It costs a few thousand dollars today, and the price is still falling. This is important because it allows us to investigate animals not usually studied in laboratories in the ecosystems and habitats in which they have evolved, which has the potential to yield more physiologically relevant data.

As a reproductive biologist, I am particularly excited that this brings us closer to understanding the great symphony — maybe cacophony — of mechanisms driving sexual behaviour and reproduction.

We recently used RNA sequencing to get a more in-depth view of how the reproductive axis in common pigeons responds to stress. Chronic stress can disrupt reproduction, and we want to know all the ways it can do this.

We are looking at the activity of every gene actively transcribed in the reproductive axis — the hypothalamus in the brain, the pituitary gland and the gonads — in response to stress. This enormous data set has produced hundreds of hypotheses to examine the effects of stress on what could be newly discovered reproductive mechanisms. These will lead us towards targets for genetic intervention or therapy for the millions of men and women who report fertility problems.

But we can also benefit from taking a step back and examining whole animals in the real world. For example, feral pigeons could serve

as powerful models for assessing what effects the exposure to hazardous substances in the environment, or the ‘expose-ome’, has on the reproductive axis. Our findings show that free-ranging pigeons experience similar exposure threats in the environment as humans living in the same neighbourhood. We can use pigeons as canaries were once used in coal mines, as bioindicators of hazardous substances in the environment. Sequencing techniques can then allow us to determine how those exposures affect the well-conserved reproductive system.

We can take our shiny new technologies and marry them with ‘old-school’ scientific tools to expand discovery in ways we never could before. We could look at pigeons in their environment in real time, characterize changes in their genomes and proteomes and see the effects on reproduction. We are the natural historians, at the level of the gene, of our time.

VIVIENNE MING

Making an Internet of Scientific Things

Theoretical neuroscientist and executive chair, Socos Labs, Berkeley, California.

The Internet of Things, all of those Internet-enabled devices that are becoming so common in homes today — Alexa, Google Home, Nest thermostats, smartphones — they are the sensors and actuators of a massive swarm intelligence. We think of a single Alexa device, the Internet-connected smart assistant developed by Amazon, as a lone personal assistant, but it is more accurate to recognize it as part of a massively distributed multisensor array extending into millions of homes and feeding

an enormous experimentation system that is the true Alexa. Rather than millions of individual robots, it is a single artificial intelligence (AI) that is constantly learning about the world, with the actions of one family influencing its exploration, and exploitation, of another.

Those distributed intelligences are transforming our lives, but they could also be transformative in the sciences. I would like for, and believe we are ready for, researchers to begin collaborating on a distributed Internet of Scientific Things (IoST) — an open system for connecting distributed sensors and actuators to a powerful machine-learning platform driving global-scale experiments. Even simple versions of this system have enormous power. Google found that its smartphones could pick up on early symptoms of Parkinson’s disease from changes in gait detected by the phone’s accelerometer and gyroscope. Using an expanded set of smartphone sensors, my team was able to predict the onset of manic episodes in people with bipolar disorder. But right now, that sort of experimental power is inaccessible to most scientists.

Imagine if researchers could access data from smartphones, smartwatches and appliances running IoST apps, along with more conventional sensors used in experiments around the world? Add to that AI systems mining for relevant published research and data already out there in your field. Similar to how current commercial AI identifies hidden business connections for salespeople, the AI of an IoST would augment the work of scientists hunting for data relevant to their fields. What if my neuroimaging software was directly plugged into an IoST platform and made data accessible in real-time, not just to my lab, but to everyone in my field and beyond? Or, logging into the platform to discover the activity of five new researchers that you should meet. Imagine that.

Admittedly, there are scary elements of these massively distributed systems. Will certain organizations have restrictive control over the data? Will findings from these new platforms go through traditional scientific publishers, through companies such as Alibaba or Amazon, or through open-access platforms like GitHub and arXiv? Serious issues of access and research ethics must be addressed, but transformation is looming.

Already, individual labs and researchers are leveraging the possibilities. But the scientific community must take the lead. If we, as scientists, build these systems ourselves, we can make publishing more egalitarian, data collection more sharable, and science more transparent. Otherwise, someone else will build it for us. But the amazing tradition that is science should not be obfuscated in the hands of just a few people. ■

INTERVIEWS BY KENDALL POWELL

These interviews have been edited for length and clarity.



Devices such as this Apple Watch are inspiring the development of an Internet of Scientific Things.

CAREERS

ECOLOGY First full survey of career and salary published for US PhD holders **p.537**

PHYSICS Lab instructors urged to give students greater freedom **p.537**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

CHELSEA GARDNER



BY TRACI WATSON

Archaeologist Louise Hitchcock went to Israel in 2017 to look for artefacts from the Iron Age. She also found something else — a large dog who wormed his way into her heart.

The mongrel was skinny and skittish when he appeared at the dig site where Hitchcock, who studies Greek prehistory at the University of Melbourne in Australia, was working. A previous owner had apparently dumped him, and no local family wanted him. “I liked him, and I just couldn’t let him be abandoned again,” Hitchcock says. Apprehensively, she shipped him home.

Today, the 35-kilogram saluki mix, who is named Fred, enjoys a life of walks and belly rubs. Although he was fearful and occasionally aggressive in Israel, he has since become mellow and friendly, Hitchcock says. “He will now go up to strangers and want to have his neck scratched.”

Like Hitchcock, many researchers spend substantial time in the field, where they might confront large numbers of apparently homeless cats and dogs. For some scientists, animals that wander up to a research site provide welcome companionship. But they can also pose a direct threat to scientific projects (see ‘Hounded and harassed’).

WILD ROVERS

Free-roaming ‘street dogs’ number perhaps 300 million globally, says Andrew Rowan, chief scientific officer at the Humane Society of the United States, based in Washington DC. Studies of ‘street cat’ numbers are limited, but Rowan’s self-described “crude estimate” of the worldwide cat total is 700 million; this includes cats that live in a community setting and those claimed by humans.

Inevitably, some of these street animals are hurt or sick, prompting some scientists to intervene or even to adopt one. Yet helping a suffering creature while in the field is not always easy. Time is often limited, and a step that would be simple at home, such as taking an injured animal to a vet, can be daunting in a region where such services are scarce.

But social media — which provides easy access to information and resources — and a growing global network of voluntary groups have eased the path for scientists who are troubled by the plight of needy animals.

The lives of free-roaming animals — and local attitudes towards them — vary ►

Peeping tom: a street cat looks on as an archaeologist examines a stone wall in Thasos, Greece.

ANIMAL PHILANTHROPY

When friendship is a whisker away

Cats and dogs that live on or around field sites can become cherished companions, but adopting them can be tricky.

► enormously. In parts of the Caribbean, street dogs are well nourished and treated affectionately, says ecologist Ryan Boyko, who heads the canine DNA-testing company Embark Veterinary in Austin, Texas, and has sampled DNA from street dogs in nearly 40 countries. But in other places, such as parts of Africa, street dogs are emaciated, riddled with open sores and covered with parasites.

The dog that ultimately found a home with Hitchcock was, by comparison, in good health. Although thin and tick-ridden, he was neither injured nor sick. His sudden appearance at the dig site, in a remote national park, led Hitchcock to suspect that someone had left him to fend for himself.

Whenever possible, scientists who want to help such animals should first try contacting local groups, says Meredith Ayan, executive director at the Society for the Prevention of Cruelty to Animals International in New York City. Facebook is a good resource for finding local rescue groups. Community members are likely to know which animals that seem to be strays are actually associated with a household and don't require feeding or care. Animal organizations with an international presence and vets might also be able to help.

Many animal-welfare groups advise caution for scientists in the field when it comes to

feeding animals. Indiscriminate handouts, say group representatives, might create friction with local people and disrupt animals' routines. "Once you leave, these animals don't have food again," says Joy Lee, who until the end of 2017 was based in Ahmedabad, India, where she worked for Humane Society International, an animal-protection organization that is active around the world.

PET PREDICAMENTS

That fate was exactly what Hitchcock feared for Fred and a second dog that appeared with him. "People were turning them into pets," she says. "I thought, 'In three weeks we're going to leave and they're just going to be discarded again.'" It seemed unlikely that Fred and his canine companion Fi belonged to the closest village, which was five to ten kilometres away.

Besides, Fred was timid, whereas the local sheepdogs were so aggressive that, for her own safety, Hitchcock feared to approach the shepherds to make enquiries. Worried about Fred and Fi's future, she and others began working to find them homes.

An archaeologist friend helped Hitchcock to contact a local professor of veterinary medicine, who gave the dogs inoculations and tick medication. An Israeli scientist working at the site used the messaging app WhatsApp to find

an adoptive family. But Fred's family decided to give him up, and his only remaining option was a job as a prison guard dog. "That didn't sound very good to me," Hitchcock says. "I just had to rescue him."

Researchers who, like Hitchcock, are troubled by the plight of such animals might want to support local vaccination drives. And many animal-welfare advocates recommend campaigns to sterilize free-roaming animals and return them to the community. But such trap-neuter-return, or TNR, programmes are controversial: modelling suggests that they can curtail populations in suitable areas, yet sustained success requires intense and long-lasting efforts, and few TNR programmes are rigorously monitored (P. S. Miller *et al.* *PLoS ONE* 9, e113553; 2014).

Meanwhile, the number of community organizations devoted to animal support is rising. In many parts of the world, the chance of finding a local partner is much higher than it was even five years ago, says conservation biologist John Boone at the Great Basin Bird Observatory in Reno, Nevada, who has studied the population dynamics of street cats.

And scientists who want to do more than donate money to support inoculation and other solutions can help in other ways. On two trips home to Toronto, Canada, from a field site in Greece, archaeologist Chelsea Gardner's checked luggage included five large kennels, each holding a dog bound for an adoptive or foster family. Non-profit groups, such as Canada's Paws Across the Water, arranged the placements, often through social media.

The groups also paid the animals' airfares, which can cost Can\$700 (US\$550). It was up to Gardner, a faculty member at Mount Allison University in Sackville, to navigate customs with two or three baggage carts stacked with dogs. "We're definitely a spectacle," says Gardner. "People say, 'Are they all yours?'"

A researcher who spends an extended period at one site can help by providing a foster home. Archaeologist Maria Liston of the University of Waterloo in Canada, who does field work in Greece, donates money to Nine Lives Greece, a local group. She also fosters cats in her Athens apartment — so many last year that "I lost track, honestly," she admits, before guessing that the number was between 15 and 20.

CAN'T HELP FALLING IN LOVE

Then there are the scientists who, like Hitchcock, fall in love with an animal. Among them is Ovee Thorat, a PhD student in conservation science and sustainability studies at the Ashoka Trust for Research in Ecology and the Environment in Bangalore, India. When a scrawny kitten began loitering at Thorat's field station in Gujarat, about 1,500 kilometres away, the scientist felt compelled to help. She fed the kitten, dewormed her and named her Billo, a Hindi word that translates roughly

HOUNDED AND HARASSED

When four-footed friends turn into foes

Free-roaming cats and dogs imperil some native species, complicating the work of scientists who study those creatures. Published studies say that dogs have contributed to the extinction of 11 vertebrate species (T. S. Doherty *et al.* *Biol. Conserv.* 210, 56–59; 2017) and cats to the extinction of at least 63 (S. R. Loss and P. P. Marra *Front. Ecol. Environ.* 15, 502–509; 2017). Dogs and cats not only eat native species; they can also transmit fatal diseases.

When Abi Vanak began studying the interactions of wild carnivores at a nature reserve in India, he set up camera traps to monitor the local jackals, foxes and jungle cats. Instead, dogs were the most frequently photographed carnivore, says Vanak, a conservation scientist at the Ashoka Trust for Research in Ecology and the Environment in Bangalore, India.

Vanak and his colleagues presented authorities with a plan to reduce dogs' impact, but met with little success. If he could do it again, he would present his team's ideas to village councils, which would help to win community support, he says.

Bonaventura Majolo, a primatologist at the University of Lincoln, UK, knew of two dog attacks on wild Barbary macaques,

Macaca sylvanus, while he was studying the endangered species in Ifrane National Park, Morocco. In each case, the monkey that was bitten soon disappeared — an almost certain sign that it had died.

Conservation biologist Joel Berger acted swiftly when he and his colleagues saw dogs chase a mother and baby takin, *Budorcas taxicolor*, a threatened goat-like herbivore. After the baby became stranded in the middle of a river at Jigme Dorji National Park in Bhutan, the researchers warmed it and gave it fluids.

But they did not see it again, nor did they see two additional calves that had become separated from their mothers by dogs that year. "We assume that their fates are not positive," says Berger, of Colorado State University in Fort Collins and the Wildlife Conservation Society in New York City.

In many countries, cultural attitudes rule out the killing of invasive predators to protect wildlife, Berger notes. He urges researchers whose study animals are at risk to work with local conservation groups and government officials. He also recommends that scientists publish studies about the impact of dogs and cats to help motivate governments to act. **T.W.**

EDUCATION

Physics labs must adapt

US undergraduate students who participate in laboratory sections of physics courses show no discernible improvement in their exam scores over students who participate in only the lecture sections. In a study published in *Physics Today*, authors analysed 9 introductory physics-lab sections taught by 7 different instructors and involving nearly 3,000 students at 3 US institutions (N. G. Holmes and C. E. Wieman *Phys. Today* **71**, 38–45; 2018). The researchers compared the midterm and final exam scores of those who took the optional lab component — designed to support student learning of lecture content — with the scores of those who did not. In follow-up interviews, students said that it was important for them to make their own decisions in lab and to reflect on them, but that they were not permitted to do so in structured lab courses. The authors suggest that lab instructors could better emphasize experimentation, decision-making and critical-thinking skills. They say that students could collaborate to design experiments to test their own hypotheses for explaining surprising phenomena.

SALARIES

Ecology pay surveyed

Most people with US doctorates in ecology work in jobs related to the discipline, and the highest mean salary across all employment sectors — academia, industry, government and non-profit — is US\$84,900, according to the first fine-scale national profile of ecology careers, published on 21 December in *Ecosphere*. Using 2013 data from the US National Science Foundation, the authors found that 9,984 people earned a US PhD in ecology between 1968 and 2011 (S. E. Hampton and S. G. Labou *Ecosphere* **8**, e02031; 2017). Of the 91% with jobs related to the field, 66.1% work at US academic institutions and about 15% in municipal, state or federal government positions. Almost 12% work for businesses or are self-employed, and 7.5% work for non-profit organizations. Government ecologists earn a mean of \$84,900 and spend about 82% of their time in research and development (R&D) — the highest level for R&D across all ecology-related jobs. Those in academia earn a mean of \$62,530; those who are self-employed or at businesses earn a mean of \$82,873; and those in the non-profit sector earn a mean of \$74,722.



Chelsea Gardner with Rex, a stray dog that she befriended on a field site in Athens.

CHELSEA GARDNER

to 'beautiful girl'. Billo returned the favour, staying close to Thorat during the weeks and months when no other researchers lived at the field station.

When Thorat finished her fieldwork, she could not take Billo with her to her shared city apartment. So she asked the field-station staff and research crew to take care of her pet. Billo became the darling of the field station, where she is fed and cosseted.

“Something came of all that time and effort. It was a nice, happy ending for at least one of those dogs.”

Moving a new pet to a distant country is trickier than passing it on to colleagues. A scientist shipping an animal from India, for example, should set aside US\$2,000 at a minimum, Lee advises, though the costs will depend on the destination and airline. Securing permits in India is complicated; the alternative is an Indian pet-shipping company that can handle the red tape.

After deciding to adopt Fred, Hitchcock turned to Terminal4Pets, an Israeli pet-travel agency, to help arrange his journey. The paperwork was minimal, but the costs were not: they came to about Aus\$7,000 (US\$5,500), most of it for Fred's flights, his six-month quarantine in an Israeli kennel and a 10-day quarantine in an Australian kennel.

In some places, taking home a furry companion is easier than it was once. A single 'pet passport' is accepted by every country in the European Union and applies to dogs, cats and ferrets travelling between member states. The United Kingdom no longer imposes a six-month quarantine on incoming cats and

dogs that have EU passports. Dogs imported to the United States are exempt from the federal requirement for a rabies vaccination if departing a country without rabies.

Adopted animals from war-torn countries can experience something akin to post-traumatic stress disorder, Ayan says. “It takes a while for the stress to wear off,” she says, and in a few cases, animals cannot adapt to their new homes. Fred takes anti-anxiety medication, although Hitchcock hopes to end the regimen within a year.

Sometimes intervention leads to a happy outcome. Fred, for example, initially feared car rides, vet visits and lifts, and Hitchcock worried that he'd chew up the house and clash with her little terrier. Now, he is outgrowing his anxieties, he doesn't chew or bark and he doesn't attack Hitchcock's other dog.

Another story that ended well began on Gardner's excavation site in Greece, when she and a friend started caring for two starving, tick-covered puppies. The closest shelter was far away and overrun with dogs, so the researchers fed the puppies, gave them flea and tick medicine and acclimatized them to humans in the hope that someone would adopt them. Wary of getting attached, the scientists called them only Brown Dog and Black Dog.

When Gardner returned to the site in 2017, she saw Brown Dog again — in a yard, wearing a collar. “That's exactly what we had hoped for,” she says. “Something came of putting in all that time and effort. It was really a nice, happy ending for at least one of those dogs.” ■

Traci Watson is associate editor of *Nature Research Highlights*.

CLOCKING OUT

Time to go.

BY PRESTON GRASSMANN

As May walked back through the time zones of her company town for the last time, she stared at the hands of her watch. She could feel her mind adjusting to its tiny motions, its auto-syncing movements slowing down as she passed through each border. Company news and notifications pulsed over her vision in a palimpsest of coded text.

It meant nothing to her now.

"What will you do on the outside?" Raja had asked her that morning.

Her over-clocked mind offered a range of possibilities, but only one of them made sense.

"Get my life back together," she had said.

Once, she was naive enough to believe that her mental enhancements would offer solutions to the problems in her personal life. After all, she had developed new patents and coded a submodule on robotic proprioception that would've taken years on the outside.

But this was only part of the life she had made for herself, and she finally understood that there were more important things to think about than patents and proprioception.

I can't live this way anymore.

She could hear those words in the distant sound of cars passing outside, in the chiming of a clock tower, in the cries of children playing in a park. In the subjective space and time of her own mind, they were still too far away to reach.

"Get your life back together?" his doubting voice had echoed, staring at her as if those words had been arcane ciphers that needed to be translated. She could see his thoughts racing, trying to unpack the question into logical solutions. "Think of what you've sacrificed to be here."

"I know what I've given up," she had said. *Something has to give.*

"Are you sick? Whatever it is, we can help you deal with it..."

It was common enough in a fast-town like hers. Some of the workers couldn't take the strain of moving between zones. During their brief company breaks, the time-zone shifts would require careful adjustments at

each level, allowing workers to acclimatize to faster or slower rates of clock-speed. Despite the precautions, there were still documented cases of permanently over-clocked minds, or workers suffering from mental time-lag, forced to live out their



lives in the company towns.

"No, I'm not sick. But I won't be coming back," she had said, simply.

"Have you forgotten what it is we're working towards here?" he had said. "It's not just about changing the working conditions of the world outside. It's about remaking the human condition itself."

She had heard it all before. The edge of a human singularity was a dangerous place to be, but it was also a privilege.

Her time zone was often referred to as the eighth and inner circle on the outside, regarded with an equal measure of fear and respect. They were seen as wraiths, living out their lives for the sole purpose of CAD and code. For those on the inside, it was known as the pinnacle of productivity, the peak of human over-clocking. And she had worked her way up, knowing the risks involved, until she had all but forgotten what it meant to live on the outside.

Maybe it would be better for the both of us if I left.

In that moment, she hadn't been able to respond, the words too abstract and far-away; slow-motion sounds escaping through a long and languorous sigh. And when the

realization had finally come that she had to make a choice, all she had wanted was to over-clock her mind, believing that it would give her a solution, a way out of her misery. But as she made her way through each time zone, she felt the space between them growing, until it seemed like a chasm she could no longer bridge. She was getting farther and farther away, each increment of enhancement putting more distance between them.

What if I leave the company?

How many times had she said that before?

Now that she was finally leaving, she had to wonder why it had taken her so long. The square buildings and the tea gardens and the company shops were full of scurrying motions. She walked by a crowd conversing in Goglot, a tiny cloud of private drones purring above them. She had spent years learning the native tongue of fast-towns, but for all its efficiency and compactness, it seemed so empty of meaning.

She fought through the time-lag, walking through the gate as carefully as she could, letting her natural proprioception take over. She closed her eyes and took off her watch, her mind slowing down, her thoughts syncing to the natural rhythms around her. She could only hope that they would mean something to her again.

"Am I late?" a thin voice said, slow and beautiful. It stretched and held itself in her mind. She opened her eyes then, and felt the time and distance between them closing.

A slow smile came to her lips as she reached out to bridge that final space and said: "You're just in time. You're just in time." ■

Preston Grassmann is a contributing editor of *Locus Magazine*, where he writes a regular column called *The Cosmic Village*. His recent work has been published in *Daily Science Fiction*, *Mythic Delirium* and *AE: Canadian Science Fiction*. He currently lives in Tokyo.

ILLUSTRATION BY JACEY